NBER WORKING PAPER SERIES

PUSHING CRIME AROUND THE CORNER? ESTIMATING EXPERIMENTAL IMPACTS OF LARGE-SCALE SECURITY INTERVENTIONS

Christopher Blattman
Donald Green
Daniel Ortega
Santiago Tobón

Working Paper 23941
http://www.nber.org/papers/w23941

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 20177

This research would not be possible without the collaboration of the National Police of Colombia and the Mayor's Office of Bogotá, in particular Bogotá's Secretary of Security, Daniel Mejía, who co-conceived the interventions and experiment. Innovations for Poverty Action in Colombia and the Center for the Study of Security and Drugs at Universidad de los Andes coordinated all research activities, survey data was collected by Sistemas Especializados de Información, and for research assistance we thank Juan Carlos Angulo, Peter Deffebach, Marta Carnelli, Daniela Collazos, Eduardo Garcia, Sofia Jaramillo, Richard M. Peck, Patryk Perkowski, Oscar Pocasangre, María Aránzazu Rodríguez Uribe, and Pablo Villar. For comments we thank Thomas Abt, Roseanna Ander, Adriana Camacho, Aaron Chalfin, Marcela Eslava, Claudio Ferraz, David Lam, Leopoldo Fergusson, Nicolás Grau, Sara Heller, Daniel Mejía, Ben Olken, Jan Pierskalla, Tristan Reed, Jacob N. Shapiro, Rodrigo Soares, Juan F. Vargas, David Weisburd, Dean Yang, and numerous conference and seminar participants. Data collection and analysis were funded by the J-PAL Governance Initiative, 3ie, the Latin American Development Bank (CAF), Fundación Probogotá, Organización Ardila Lülle through its funding for the Center for the Study of Security and Drugs at Universidad de los Andes, the Administrative Department for Science, Technology and Innovation of the National Government of Colombia (COLCIENCIAS), and the J. William Fulbright Program. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Pushing Crime Around the Corner? Estimating Experimental Impacts of Large-Scale Security Interventions
Christopher Blattman, Donald Green, Daniel Ortega, and Santiago Tobón
NBER Working Paper No. 23941
October 2017
JEL No. C93,K42,O10

ABSTRACT

Bogotá intensified state presence to make high-crime streets safer. We show that spillovers outweighed direct effects on security. We randomly assigned 1,919 “hot spot” streets to eight months of doubled policing, increased municipal services, both, or neither. Spillovers in dense networks cause “fuzzy clustering,” and we show valid hypothesis testing requires randomization inference. State presence improved security on hot spots. But data from all streets suggest that intensive policing pushed property crime around the corner, with ambiguous impacts on violent crime. Municipal services had positive but imprecise spillovers. These results contrast with prior studies concluding policing has positive spillovers.

Christopher Blattman
Harris School of Public Policy
The University of Chicago
1155 E 60th St.
Chicago, IL 60637
and NBER
blattman@uchicago.edu

Donald Green
Department of Political Science
Columbia University
420 W. 118 Street
7th Floor IAB
New York, NY 10027
dpg2110@columbia.edu

Daniel Ortega
Banco de Desarrollo de America Latina (CAF)
& Instituto de Estudios Superiores
de Administracion
Centro de Politicas Publicas
Ave. IESA, Edif. IESA
San Bernardino, Caracas
Venezuela 1010
dortega@caf.com

Santiago Tobón
University of Chicago
Pearson Institute and Harris Public Policy
& Innovations for Poverty Action
Peace and Recovery Program
Calle 98 No. 22-64 Of 307
Bogotá, Colombia
tobon@uchicago.edu

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/1156

1 Introduction

Police and city workers are the everyday face of the state. These street-level bureaucrats provide the most basic public goods we expect from government, especially security. The response to a crime, the picking up of garbage, and the lighting of streets—it is impossible not to notice when they are done poorly. When crime and violence start to get out of control, these are also the first levers that governments pull. Cities step up enforcement, they put more police on the streets, or they light up or clean up high-crime places.

In the United States, more than 90% of police agencies use some form of "hot spots policing," intensifying police attention to high-crime areas.[1] These tactics typically target units as small as a street segment or a specific corner. Some cities also change the quality of policing in hot spots, enforcing minor infractions with a "zero tolerance" approach.

Another tactic is to reduce disorder in hot spots through municipal services. Services can make it more difficult to commit crimes, by lighting dark areas or increasing the amount of people on the street.[2] Services may also signal order and state presence, telling criminals to stay away and citizens that the state is present. Altogether, policing and services interventions grow out of the famous "broken windows" hypothesis.[3]

This is state building on a different margin than in weaker states, but it uses the same tools and rationale. From Afghanistan to Iraq or the Philippines, militaries use security forces and public services to establish order and legitimacy.[4] In more stable places, like Bogotá, the state already has some control and legitimacy on most city streets. They are increasing state presence on the intensive margin—the last mile of state building.

This raises a number of questions. How much can more state presence reduce crime and violence? Which levers are most effective? Are there increasing or decreasing returns to state presence? Perhaps the most important but difficult question raised, however, is whether targeted state presence reduces overall crime, or merely displaces it elsewhere.

We tackle these questions in Bogotá, the capital of Colombia. Two percent of the city's 136,984 streets accounted for all murders and a quarter of all crimes from 2012–15. These "hot spots" received less than 10% of police time and limited public services. In January

[1] Weisburd and Telep (2016); Police Executive Research Forum, (2008). Interventions include greater police time, greater traffic enforcement, aggressive enforcement of infractions, and problem-oriented policing.

[2] Police presence and street lighting are meant to raise the risk of detection and capture for offenders—a tenet of the economic approach to crime prevention where crime is a gamble and increasing expectations of apprehension and punishment deters people from crime (Becker, 1968).

[3] Wilson and Kelling (1982); Apel (2013). "Broken windows policing" can mean intensive, zero tolerance policing. But more visible state presence and physical order should send similar signals.

[4] Besides fighting insurgents, intensifying security and public services are designed to win the "hearts and minds" of citizens. The idea is that they will be more likely to inform on offenders or collaborate against the insurgents. See Berman and Matanock (2015) for a review.

2016, a new city government decided to try increasing state presence in hot spots. They wanted to improve security and raise citizens' trust in police and local government.

We worked with the police to identify an experimental sample of 1,919 hot spot street segments. A segment is a length of street between two intersections, a common unit of police attention (Weisburd et al., 2012). The city first doubled police patrol time on 756 segments. Then they targeted 201 segments for clean-up and better lighting. We randomized assignment to intensive policing, more municipal services, both, or neither.

The city modeled its interventions on standard U.S. practices and evidence. Like Bogotá, crime in large U.S. cities concentrates in a small number of hot spots. Based on several experimental trials, there is a consensus in the U.S. that targeting hot spots with more state presence reduces crime within treated areas.[5] The enthusiasm for intensive policing is bolstered by two systematic reviews that argue that the evidence points to reductions in crime in nearby streets (Braga et al., 2012; Weisburd and Telep, 2016).[6]

Spillovers and the aggregate effects on crime are difficult to pinpoint, however, because of the small size of most studies.[7] The median study in existing reviews has fewer than 30 treated hot spots per treatment arm, and the largest has 104. These sample sizes make it difficult to detect large effects, even those as large as 0.4 or 0.5 standard deviations in size (see Appendix A). As a result, these studies cannot rule out huge spillovers in either direction. Given the scale of Bogotá's experiment, however, this study can identify direct effects of 0.15 standard deviations and spillovers as small as 0.02 standard deviations.

Latin America is an important place to study the state's crime fighting abilities. It is the most violent region in the world, with 42 of the 50 most dangerous cities and a third of the world's homicides.[8] Major cities also have fewer police per person than the U.S. or Europe.

[5]Chalfin and McCrary (2017) review the evidence on increased policing and find that more police are usually associated with falling crime city-wide. Looking at targeted hot spots interventions, a systematic review of hot spot policing identified 19 eligible studies (including 9 experiments). Among 25 tests of the core hypothesis, 20 report improvements in crime (Braga et al., 2012). These evaluations are largely in the U.S. Exceptions include quasi-experimental studies such as Di Tella and Schargrodsky (2004) in Buenos Aires, and ongoing experimental evaluations in Medellin (Collazos et al., 2017) and Trinidad and Tobago (Sherman et al., 2014). The evidence on interventions that tackle disorder is limited. Braga et al. (1999) and Braga and Bond (2008) report significant reductions in crime following a combined treatment of intensive arrests and environmental interventions in small U.S. cities. There is some evidence that street lighting reduces crime (Farrington and Welsh, 2008). Cassidy et al. (2014) review five studies suggesting there is weak evidence that urban renewal reduces youth violence.

[6]Banerjee et al. (2017) see displacement from drunk driving checkpoints in India. We consider this an important but distinct phenomenon from property and violent crime.

[7]Beyond methodological difficulties, prior studies have been designed mainly to address direct treatment effects and study spillovers as a secondary outcome. One exception is Weisburd et al. (2005), who study drug and prostitution hot spots. Their findings suggest the benefits from the intervention diffuse to nearby areas.

[8]See Consejo Ciudadano para la Seguridad Pública y Justicia Penal and Global Study on Homicide 2013.

Policymakers are interested in the returns to higher quality or quantity of policing.

In Bogotá, the Mayor's office first reallocated existing police patrols to spend more time on high crime streets. No new police were added in the city. Within their patrol area (a quadrant), officers were told to double their time on two hot spots from roughly one to two hours a day, in multiple visits. This intensive policing lasted from February to October 2016. With an average of 130 segments per quadrant, there was little effect on patrol time on other segments. Patrols simply went about their normal duties, interacting with citizens, and stopping and frisking suspicious people. Shortly afterward the city decided to tackle social disorder by repairing lights and cleaning up trash.

We designed the study to measure spillovers flexibly. Treating one hot spot can affect the outcomes of control hot spots. For example, criminals may shift activities to nearby hot spots, and places close to treated segments have to be crossed to deliver interventions. Thus spillovers pose an identification problem for direct effects. We are also interested in spillovers to nearby streets outside the experimental sample, or "non-hot spots." Taken together, these two spillovers tell us whether crimes are deterred or pushed around the corner.

Since we don't know the structure of spillovers, we pre-specified a more flexible design over many possible catchment areas. We divided control hot spots into categories: 0–250 meters (m) from a treated hot spot; 250–500m; and >500m. By comparing outcomes across treatment and control categories, we can first test for spillovers in the 0–250m and 250–500m regions, and then use unaffected regions as a control group for estimating the effects of direct treatment. We estimate spillovers into the non-experimental sample the same way.

Spillovers present other estimation challenges, however. By simulating the experiment many times, we show that the close proximity of hot spots leads to hard-to-model patterns of clustering, also known as "fuzzy clustering" (Abadie et al., 2016). In most randomizations, hot spots close to other hot spots tend to be assigned to spillover status. This biases treatment effects and understates standard errors. Without a fixed geographic unit of clustering, we cannot use standard correction procedures. This is a common but relatively underexplored problem with experiments in dense social or spatial networks. We show that randomization inference provides exact p-values in such settings.

To evaluate impacts, we first look at police administrative data on reported crimes. Police data are problematic, however, if errors in crime reporting are correlated with treatment. Thus we also conducted a survey of about 24,000 citizens. The survey measured unreported crimes, perceptions of security, and attitudes toward the state. Besides providing new outcomes, these data help us test whether official crime reporting is correlated with treatment.

Broadly speaking, we find that increasing state presence deters crime on treated streets, but that intensive policing pushes crime into nearby segments. In directly treated hot spots, both forms of state presence reduce crime and people's perceived security risks by more than 0.1 standard deviations. These impacts are statistically significant when we ignore spillovers. When we account for interference between units, however, the direct effects on perceived security are smaller and less precise. Even the most generous estimates, however, point to modest total effects. Fewer than 100 crimes were deterred across all treated hot spots over eight months.

The crime impacts were greatest in the 75 hot spots that received both interventions. In this case, crime and perceived security risks fall by more than 0.3 standard deviations. The difference between getting both and one treatment is not statistically significant, but it points in the direction of increasing returns to state presence on these streets.

We do not find evidence that improving state presence increased trust in the state. If anything, intensive policing reduced people's opinions of the Mayor's office. It is difficult to say why. Possibly, intense police presence intimidates or upsets some residents, although our qualitative investigations show no evidence of this.

Meanwhile, we see evidence that intensive policing pushed crime around the corner. We look at the sample of 77,000 non-hot spot segments within 250m of the experimental sample. Being close to an intensively policed hot spot increases reported crimes. In total these spillovers more than offset the direct treatment effects. While imprecise, these estimates allow us to rule out a sizable aggregate fall in crime.

In our main specification, it is mainly property crime, as opposed to violent crime, that gets pushed around the corner. There is weak evidence that the interventions led to a decrease of nearly 100 homicides and sexual assaults—an 8% decline. This difference between violent and property crime is statistically significant. This result seems to be sensitive to specification, however. Violent crimes do not decrease across all specifications. Thus the distinction must be taken with caution.

On the positive side, the benefits from municipal services may diffuse to nearby streets. These estimates are imprecise and must also be taken with caution. When we assume that crime spillovers follow an exponential rate of decay, however, the positive spillovers resulting from the municipal services intervention are significant at almost the 10% level.

These results show the importance of small spillovers and statistical power of the experiment. The cumulative effect of many tiny spillovers is obviously important in evaluating the interventions and understanding the relationship between state presence and violence. This is especially true when we need to assess the aggregate effects on crime, or distinguish between types of crime. Even with a sample size that is an order of magnitude greater than previous experiments, the spillover and aggregate effects are difficult to identify. Thus, methodologically, this study illustrates the importance of scale in estimating the effects of place-based interventions, and the importance of accounting for interference between treatment and control units. It also shows the importance of using randomization inference to avoid overstating precision.

Our results, if true more generally, also add some nuance to a common argument in criminology: that crime and violence are concentrated in a small number of people, places, and behaviors; and that targeted interventions stand the best chance of being effective.[9] Alongside another large-sample study of policing, of drunk driving checkpoints by Banerjee et al. (2017), our evidence reinforces the idea that crime is concentrated, but targeting places may not be generally effective as crime may simply be pushed around the corner.[10] If place-based interventions simply displace crime, then targeting high-risk people and behaviors could be more impactful to address this kind of criminal behavior.

There are parallels between our results and the historical literature on states, where the most common response to state coercion has been for people to elude the state or run away (Scott, 2014). The perennial problem of state building is controlling people, not land. The evidence from Bogotá suggests it could hold true even in the last mile of state building.

2 Setting

Bogotá, a city of roughly 8 million people, is the industrial and political center of Colombia. In 2015, Bogotá's GDP per capita was \$9,612 at market exchange rates, or about \$22,000 adjusted for purchasing power parity (PPP). 10% of the population was below the national poverty line for metropolitan areas of PPP\$6 a day, and 2% was below the extreme poverty line for metro areas of PPP\$2.50 a day. Many poor were displaced by a low-intensity civil war that ran for a half century until a 2016 peace agreement.

2.1 Crime and policing in Bogotá

Crime is one of the most pressing social problems in Bogotá. In the 1990s Bogotá was one of the most violent cities in the world, with 81 murders per 100,000 people.[11] In 2016 the figure was 15.6. This is much lower than the most violent cities in the world, such as 120 in Caracas, 65 in Cape Town, 64 in Detroit, and 64 in Cali, Colombia. It is comparable in crime rates to a U.S. city like Chicago, with 15 murders per 100,000 in 2015, but greater

[9] Braga et al. (2012); Abt and Winship (2016); Weisburd and Telep (2016); Weisburd et al. (2017)

[10] Similarly, Blanes i Vidal and Mastrobuoni (2017) use natural, high-frequency variation in police presence in the U.K. to argue that the deterrence effect of police lasts for a maximum of 30 minutes.

[11] It had 81 murders per 100,000 people in 1993. A number of factors are said to have contributed to the improvement, including the decline in civil war, as well as advances in police capacity, gun control policies, restrictions on alcohol consumption, and a major local security push.

than the 7 recorded in Los Angeles or 4 in New York.[12] As in cities like Chicago, despite improvements crime remains one of the foremost social and political concerns.

The nature of Bogotá's crime varies, from pickpocketing and cell phone theft in busy commercial areas, to burglary of businesses and homes, to drug sales and any resulting violence. Most violent crimes are of passion. The Mayor's office estimates that 81% of all the homicides in the city in 2015 were a result of fights, 12% were contract killings, and 5% from violent robberies. Most offenders are individual young people. There are some semi-organized youth gangs, and some organized crime, but they do not seem to be responsible for the vast majority of the street crime or violence.

Like many cities, crime in Bogotá is also highly concentrated. According to official crime statistics, from 2012 to 2015 just 2% of the city's 136,984 street segments accounted for all murders as well as a quarter of all other reported crimes. These hot spots are distributed around the city. They include wealthy areas where criminals come to mug pedestrians, burgle homes, or steal expensive cars, as well as more barren industrial areas with little traffic, where it is easier to sell drugs or steal. Hot spots also include popular nightlife areas.

Security policy and policing Bogotá has moderate to low levels of police compared to large U.S. and Latin American cities. Bogotá has about 18,000 police officers in operational activities, including about 6,200 patrol agents. We estimate about 239 police per 10,000 people. The Colombian average is 350, and most cities are above Bogotá's ratio. The national U.S. ratio in the U.S. was 230 in 2013 but is greater in large cities, including 413 in New York, 444 in Chicago, 611 in Washington, or 257 in Los Angeles.[13]

Patrols are instructed to spend more time in high-crime places but do not necessarily comply. One indication is that 2% of streets account for a quarter of all crime, but we estimate they received roughly 10% of police patrol time during 2012--15.

The police freely patrol almost all city streets. Patrols are reasonably well-regarded. The broader police force is not without problems, but our citizen survey (detailed below) suggests that street patrol officers are regarded as competent and non-corrupt.

In January 2016 a new Mayor came to power, Enrique Peñalosa. Crime reduction and increasing trust in government were central to his platform. In his first 100 days, the Mayor pledged to dedicate more municipal services and law enforcement in 750 hot spots.

Municipal services included trash collection, tree pruning, graffiti clean-up, and streetlight maintenance. The performing agencies report directly to the Mayor's office, but the Mayor's power is limited by contracts and difficulties monitoring and enforcing instructions.

[12]U.S. figures come from the FBI Uniform Crime Report and others from the World Atlas.

[13]Data for Colombia was reported by the Secretariat of Security of Bogota, data for the U.S. is from the Department of Justice Statistics, and other data is from the United Nations Office on Drugs and Crime.

When it comes to the police, the Mayor's office can influence tactics, force allocations, and equipment, but has little say in total force size. City police forces in Colombia are a branch of the National Police and report up to the Minister of Defense. But the city has the power of the purse, as it pays for police equipment. The Colombian Constitution also calls on police to comply with the Mayors' requests and policies. Changes in force levels are much more expensive, however, and the national government rejected the Mayor's request to increase the number of police. Thus the Mayor's office focused on increasing police efficiency and quality, especially street patrols.

Patrolling The quadrant (*cuadrante*) is the basic patrolling unit. Bogotá has 19 urban police stations. Stations are divided into CAIs—*Comando de Atención Inmediata*—a small local police base that coordinates patrol agents and takes civilian calls. Each CAI has about 10 quadrants. There are 1,051 quadrants, with 130 street segments on average.

Each quadrant has six permanent patrol officers. They patrol in pairs, on motorbike and foot, in three shifts of eight hours each. In practice, patrols are expected to move about throughout their shift, by motorbike. They may patrol a street on motorbike or dismount to speak to shopkeepers, passersby, and suspicious people.

Patrols carry a handheld computer that allows them to check a person's identification number for outstanding warrants. Patrols have daily quotas. They are expected to regularly stop and frisk any suspicious people, and will seize illegal weapons (usually knives) and other contraband. Patrols tend to focus interrogations on young men. An arrest means both patrollers must take the suspect to the station, for hours of paperwork and processing. This keeps them from meeting performance goals, and so patrols may avoid minor arrests.

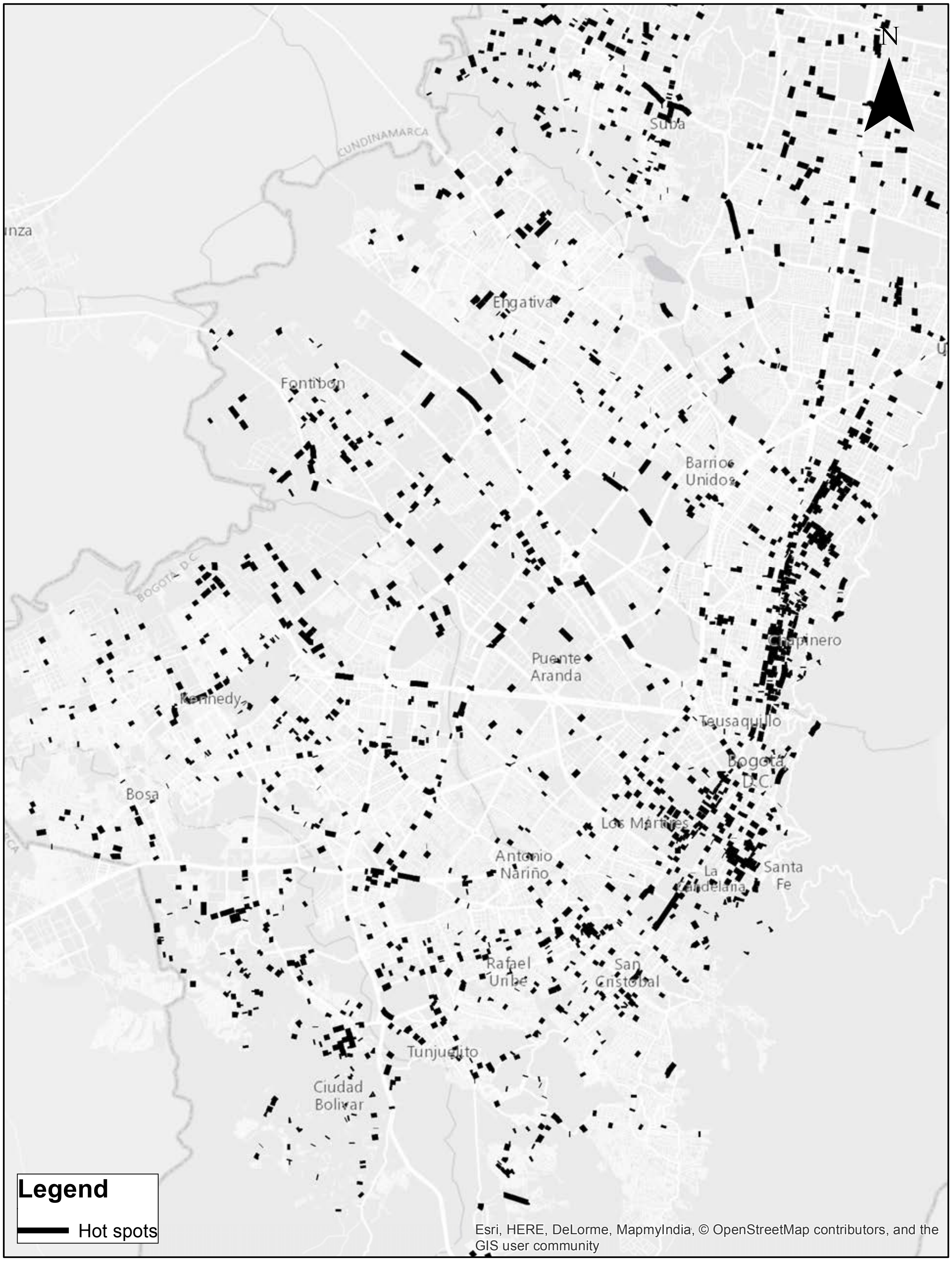
The handheld computer also contains a global positioning system (GPS) chip that records the patrol's location roughly every 30 seconds (when operational). The city first piloted and introduced the system in late 2015, under the previous Mayor. The new system lets station commanders view patrol positions in real time and get regular performance statistics. Thus the study period is a period of increased monitoring and measurement of patrol activity.

2.2 Hot spot identification and the experimental sample

Figure 1 maps Bogotá's 136,984 street segments and indicates the 1,919 hot spots in our experimental sample. To create this sample, we started with the 2% highest-crime segments, using an index of reported crimes from geo-coded official statistics, between January 2012 and September 2015.[14] We then asked each station's commanders and staff to verify the hot

[14]We constructed a geo-fence of 40m around each segment and assigned a reported crime to that segment whenever it fell within its geo-fence. Appendix B reports further details. A calculation error meant that

Figure 1: Map of hot spots



Notes: Hot spot street segments, in black, are the 1,919 streets included in our experimental sample.

spots. Official crime data omit most petty crimes and disorder. Calls or informal reports to police do not show up in official statistics, and police do not record crimes they observe. Based on their knowledge, the police eliminated about a third of the hot spots, adding others in their stead, leaving 1,919 segments that account for 21% of the city's reported crimes.[15]

Table 1 reports summary statistics. In October 2016, the police updated all 2012–16 crime data with more accurate GPS coordinates and additional crime categories, and we report original and updated data.[16] Hot spots had between 0 and 82 crimes reported in the previous four years (461 with the updated data as we had information on more crime types), with an average of 5 crimes.[17] More than half were property crimes, but violent crimes such as murders and assaults were also important. 95% of hot spots had relatively low levels of physical disorder such as garbage.

3 Interventions

Intensive policing Intensive policing began on February 9, 2016 and ended on October 14, 2016.[18] Intensive policing generally meant a two-thirds increase in police patrol time. As we will see below, during the intervention control streets received roughly 92 minutes of patrol time on average, with treated streets receiving an additional 77 minutes—an 84% increase.[19] In order not to overextend patrols, the police required us to assign no more than two hot spots to treatment per quadrant so as not to distort regular duties too much. A 77-minute increase on two hot spots implied that patrol time fell on other segments in the quadrant by roughly one minute each.

608 segments outside the top 2% were included in this initial sample. These were generally high crime segments, as 90% of those streets were above the 95th percentile of baseline crime, and all were above the 75th percentile. In retrospect, this error proved useful since it gave us more variation in baseline crime levels, which we use to study treatment heterogeneity.

[15]Homicides are recorded by police. For any other crime to be included in the database, victims had to travel to one of 19 police stations, file a formal report, and include relevant details such as location. Our endline survey (discussed below) suggests that official statistics record only about a fifth of all crimes.

[16]Some crimes moved to nearby segments, and the correlation between the old and new data is 0.35 at the segment level and 0.86 at the quadrant level. These corrections were unrelated to this study.

[17]Quadrants with at least one hot spot had an average of 3.5 reported crimes per segment across the whole quadrant, while the average quadrant in the whole city reported 1.5 crimes.

[18]The government, however, did not publicize the eligible high-crime streets, the existence of an experimental design, or which specific streets were being targeted. The Mayor's office initially planned to run this intensive policing intervention for at least 4 to 6 months. They extended the intervention in part to permit the research team enough time to fund and conduct a survey of citizens.

[19]Before the intervention, 1–2 weeks of GPS data suggested that hot spots received at least 38 minutes of patrol time per day. It is doubtful that actual time rose from 38 to 86 minutes. Rather, the 38 minutes was probably an understatement of average patrolling time per hot spot. The police did not have data on pre-intervention patrol times, since the handheld computers with GPS chips were piloted November 2015 through January 2016. See Appendix B.

Table 1: Descriptive statistics for the experimental sample (N=1,919) and tests of balance (treatment versus all control streets, including potential spillover streets)

					WLS test of balance			
	Summary statistics				Intensive policing		Municipal services	
	Mean	Std. Dev.	Min.	Max.	Coeff.	p-val	Coeff.	p-val
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
# of reported crimes on street, 2012-15 (original)	4.53	5.72	0	82	-0.17	0.62	-0.13	0.70
# of violent crimes	1.88	2.94	0	56	-0.18	0.21	-0.05	0.75
# of property crimes	2.66	3.97	0	50	0.02	0.95	-0.08	0.76
# of reported crimes on street, 2012-15 (updated 10/2016)	5.18	18.24	0	461	-0.21	0.86	-0.36	0.79
# of violent crimes	1.40	5.38	0	78	0.39	0.38	0.22	0.68
# of property crimes	3.78	14.09	0	407	-0.60	0.45	-0.58	0.52
Average # of reported crimes per segment in quadrant, 2012-15	3.56	5.13	0	61	-0.30	0.50	0.38	0.49
Daily average patrolling time (11/2015 – 01/2016), minutes	38.03	70.27	1	1029	-1.77	0.73	3.42	0.57
Rating of baseline disorder (0–5)	1.18	0.74	0	5	-0.05	0.31	0.35	0.00
Eligible for municipal services	0.86	0.35	0	1	-0.02	0.27	0.22	0.00
Meters from police station or CAI	551.37	351.46	6	2805	-26.18	0.26	-11.95	0.64
Zoned for industry/commerce	0.38	0.49	0	1	-0.09	0.01	0.05	0.16
Zoned for service sector	0.13	0.34	0	1	0.02	0.33	0.03	0.25
High income street segment	0.07	0.25	0	1	0.00	0.79	-0.01	0.54
Medium income street segment	0.55	0.50	0	1	-0.06	0.06	0.00	0.98
# of segments in quadrant	127.21	86.99	2	672	2.05	0.71	-3.04	0.57
# of hot spots in quadrant	3.67	2.68	1	14	-0.30	0.08	-0.16	0.31
# segments treated with policing in quadrant	1.15	0.95	0	3	1.35	0.00	-0.01	0.91
# segments treated with services in quadrant	0.66	0.69	0	3	-0.08	0.06	0.91	0.00
Assigned to intensive policing	0.48	0.50	0	1	1.00	-	0.00	-
<250m from intensive policing	0.29	0.46	0	1	-0.56	0.00	0.01	0.83
250–500m from intensive policing	0.14	0.35	0	1	-0.28	0.00	0.00	0.96
>500m from intensive poling	0.09	0.28	0	1	-0.17	0.00	-0.01	0.72
Assigned to municipal services	0.41	0.49	0	1	0.00	-	1.00	-
<250m from municipal services	0.19	0.39	0	1	0.05	0.01	-0.31	0.00
250–500m from municipal services	0.17	0.37	0	1	-0.01	0.71	-0.28	0.00
>500m from municipal services	0.23	0.42	0	1	-0.04	0.03	-0.40	0.00

Notes: Columns 1–4 display the summary statistics for our sample of 1,919 hotspots, weighted by the probability of being in the observed experimental condition. In columns 5–8, we perform a balance test for treated vs all control units using weighted least squares.

Commanders told patrols to visit treatment hot spots at least 6 times per day for roughly 15 minutes each, mostly during the day unless near a bar. The police generally did not know what hot spots were in the control group, but in principle they could make reliable guesses. Commanders instructed patrols to continue their normal duties in treated hot spots: running criminal record checks; stopping, questioning, and frisking suspicious people; door-to-door visits to the community; conducting arrests or drug seizures; and so forth.[20]

Municipal services One city office coordinates street light maintenance and a second office is in charge of all clean-up activities. Both offices contract private companies to service the streets. Contractors were expected to perform their usual duties, but the Mayor's office gave contractors lists of segments where they were asked to assess issues and deliver the appropriate services. The municipal services intervention began April 11, 2016 and continued until the end of the intensive policing intervention.

How do the Bogotá interventions compare to other hot spots interventions? The Bogotá intervention is broadly similar in style and approach to U.S. interventions that intensify patrol time but maintain normal duties.[21] It is hard to compare since many of the U.S. studies do not describe levels of control group policing or the intensity of treatment. Given the size of its experiment, the intensity of treatment in Bogotá is probably lower than in U.S. studies. Three experiments, for instance, seem to report more intensive treatments.[22] The closest intervention in treatment intensity and size, although still at a smaller scale, is the Medellín hot spots policing program (Collazos et al., 2017).

4 Data

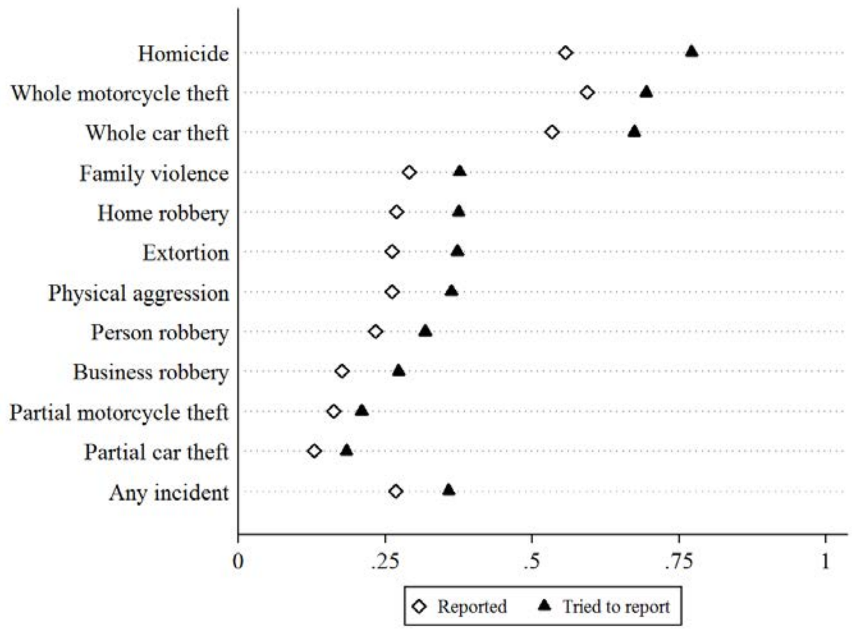
Bogotá has rich administrative data, but these sources have no information on certain outcomes of interest (such as state legitimacy), and the crime data were of questionable com-

[20]The only exception was in three streets known as "The Bronx." Early in our intervention period, the police and city invaded and cleared the three streets. This was a much more intensive, one-time intervention. Two of the three streets happened to be assigned to treatment and one had been assigned to the control group. Police cleared the streets and the city demolished the buildings. In this extreme case, it is obvious that more policing can reduce crime.

[21]Other approaches vary. Some interventions take a "zero tolerance" approach, enforcing the most minor infractions. Others focus on "problem-oriented policing," where officers try to proactively address underlying problems. Our results are not comparable to zero tolerance or problem-oriented tactics. Appendix A describes these various studies in detail.

[22]In the Minneapolis Hot Spots experiment patrol times were about 2.5 larger in treated hot spots than controls (Sherman and Weisburd, 1995). In the Philadelphia policing tactics experiment patrol times were about 8 hours/day in treated hot spots (Groff et al., 2015), and in the intensive patrolling intervention of the Jacksonville policing experiment patrols were 53 hours/week in treated hot spots (Taylor et al., 2011).

Figure 2: Proportion of crime reported, by crime (survey-based)



Notes: The figure includes data on all street segments surveyed. Each observation is a survey. The white diamonds denote the proportion of people that effectively reported a crime out of all victims. The black triangles denote the proportion of people that tried to report a crime out of all victims.

pleteness (including a danger of measurement error correlated with treatment status). As a result, we complement the administrative data with primary data collection. In the end we draw on six main sources of data prior to, during, and at the conclusion of the interventions.

1. Administrative data on police and municipal services compliance. The police shared the full database of GPS patrol locations for all 136,984 streets, 2015–17.[23] City agencies also shared reports on their diagnosis of each street and compliance with treatment for all streets assigned to the municipal services treatment.

2. Crime and policing. Police shared data on reported crimes and operations 2012–17, geolocated to 136,984 streets.[24] Many U.S. studies also use emergency call data since they are less prone to manipulation, but these data were not available.

[23]Not all handheld computers were functional at all times, and at times over 2016 the system went offline for a few days to a few weeks, and so we use data only during those periods when the system was generally operational in a given police station—on average 33 of the 37 weeks of the intervention.

[24]Prior to the intervention, we received the 2012–2015 data on the city's priority crimes: homicides, assaults, robberies, and car and motorbike theft. 77% of the crimes had exact coordinates and the rest had the address, which we geolocated ourselves, with about 71% success (or 93% of all reported crimes). We also received all data on arrests; gun, drugs and merchandise seizures; and stolen cars and motorbikes recovered. In October 2016 the police provided updated data that corrected for geolocation problems (thus retrospectively changing pre-intervention data). With the new information we also received data on reported cases of burglary, shoplifting, sexual assaults, family violence, threats, extortion and kidnapping.

3. Survey of Bogotá residents. In October 2016 we surveyed 24,000 citizens on 2,399 segments—the 1,919 in the experimental sample, plus a representative sample of 480 segments outside the experimental sample. We interviewed a convenience sample of 10 people per segment, and averaged responses over each segment. The survey collected outcomes such as: perceptions of security risks; perceived incidence of crimes; crimes personally experienced; crime reporting; and trust in and perceived legitimacy of the police and the Mayor's office. Figure 2 illustrates the difference between actual and officially-reported crimes. We asked whether or not people had experienced a crime since the beginning of the year, whether they had attempted to report it, and if they were successful. Homicides are reported by police if individuals did not report them, so administrative data probably capture most murders. But for all other crimes, about 27% of the people say they reported the crime, and an additional 9% of people say they attempted to report the crime but were unsuccessful. Reporting rates are highest for vehicle theft, because insurance claims require a report.

4. Survey of street disorder. As discussed below in section 5.2, to measure levels of street disorder before and after treatment we sent enumerators to take photographs and rate the presence of graffiti, garbage, and boarded-up buildings on a 0–5 scale.[25]

5. Administrative data on pre-treatment street characteristics. The city also shared data on pre-treatment street characteristics: urban density, income level (high, medium, low), economic use (housing, services, industry), presence of public surveillance cameras, and distance to the closest police station, commercial area, school, religious center, health center, transport station, or other public services as justice.

6. Qualitative interviews. We began with informal qualitative interviews with dozens of police officers and citizens about their experiences with the intervention and police tactics in general. We also hired observers to discreetly visit 100 streets in the experimental sample for a day and passively observe police behavior. They also interviewed citizens in each segment about police behavior and attitudes.

[25] We visited 1,534 of a total of 1,919 scheduled streets in March (three months before the municipal services intervention began) in order to narrow down the number of eligible hot spots. We did not collect data in the remaining 385 streets because of security concerns from the enumerators. (Note that there was no association between intensive policing treatment and these security concerns.) As we discuss in section 5.2, 1,459 were eligible for the municipal services interventions and 414 of them were assigned to treatment. Those streets were split in two batches of 201 and 213 streets respectively in order to randomize timing, but only the first batch was effectively treated. Then, in order to assess the levels of compliance, we sent enumerators to the 414 streets in the first and second batches in June (one to two weeks after municipal services started to be delivered) and December (two months after the end of the intervention). Again, because of security concerns of the enumerators, we visited 409 in June and 410 in December.

To simplify our analysis and deal with the problem of multiple comparisons, our pre-analysis plan distinguished primary from secondary outcomes, and pooled like measures into summary indices to reduce the number of hypotheses tested (following Kling et al. 2007).

Our primary outcomes are two insecurity measures: perceived risk and crime incidence. Table 2 reports summary statistics on a standardized index of each outcome for each of the 4×5 experimental conditions, using inverse probability weights for assignment into each of the treatment conditions. We discuss secondary outcomes, particularly the perceived legitimacy of the police and local government, in Section 6.4 below.

1. Perceived risk of crime and violence on the segment. Our citizen survey asked respondents to rate perceived risk on a 4-point scale from "very unsafe" to "very safe" in five situations, such as: for a young woman to walk alone after dark on this street; for someone to talk on their smartphone on this street; for a young man to walk alone after dark on this street; and simply the perceived risk of crime "during the day" and "at dusk". We construct a index of perceived risk that takes the average across all respondents in the segment. All indexes in the paper are standardized to have mean zero and unit standard deviation.

2. Crime incidence on the segment. We construct a standardized index of crime that equally weight the survey and administrative data. The two components include: (i) survey respondents' opinion of the incidence of crime on that segment, as well as personal victimization on that segment since the beginning of the year; and, (ii) the total number of crime incidents on that segment reported in the administrative crime data since the beginning of the intervention. We can subdivide all measures into property and violent crimes, although our main measure pools all crimes into one index.

   The survey measured perceived incidence and personal victimization by walking respondents through a list of 11 criminal activities. After finding out whether any of these activities happened on the street since the beginning of the year, we asked respondents about each crime to establish perceived frequency (ranging from "everyday" to "never" on a 0-6 scale), and whether it happened to the respondent him or herself on that segment. We show results for the two individual components in order to give a sense of the absolute impacts and differences between survey and administrative data.

Table 2: Summary statistics for the primary security outcomes, all experimental conditions

			Municipal services assignment				
			Treated	<250m	250-500m	>500m	Ineligible
			(1)	(2)	(3)	(4)	(5)
A: Perceived risk (z-score)							
Intensive policing assignment	Treated	Mean	-0.073	0.430	0.138	-0.013	-0.373
		SD	0.876	1.017	0.864	0.943	0.934
		N	75	154	150	201	174
	<250m	Mean	0.168	0.335	0.223	0.160	-0.124
		SD	1.061	1.005	0.859	1.369	1.013
		N	74	213	130	125	162
	250-500m	Mean	-0.105	0.291	0.057	0.256	-0.337
		SD	1.042	0.883	0.938	0.942	0.974
		N	32	32	75	80	75
	>500m	Mean	-0.174	0.320	0.124	-0.218	-0.651
		SD	0.914	1.078	1.042	0.912	0.994
		N	20	14	13	68	49
B: Crime incidence (z-score)							
Intensive policing assignment	Treated	Mean	-0.079	0.379	-0.056	-0.047	-0.179
		SD	0.808	1.010	0.790	0.868	0.877
		N	75	154	150	201	174
	<250m	Mean	0.157	0.425	0.139	0.169	0.248
		SD	1.032	1.056	0.849	1.769	1.230
		N	74	213	130	125	162
	250-500m	Mean	-0.143	0.207	-0.053	0.096	-0.105
		SD	0.825	1.024	0.889	0.921	0.874
		N	32	32	75	80	75
	>500m	Mean	-0.215	0.361	-0.147	-0.325	-0.419
		SD	1.092	1.297	1.024	0.745	0.862
		N	20	14	13	68	49

Notes: We report weighted means for each experimental condition, where weights are the inverse of the probability of falling in the corresponding treatment condition. We estimate that probability with repeated simulations of the randomization procedure. The ineligible condition in Column 5 reflects those streets that did not exhibit any disorder at baseline. Technically there are 3×4 ineligible conditions for each dependent variable, one for each relative distance from municipal services treated streets, but we pool those columns here for simplicity.

5 Methodology

The size and direction of spillovers drive the policy implications of place-based anti-crime programs. Failing to account for spillovers could also bias our estimates of direct treatment effects. If control hot spots are close enough to treated hot spots to experience displacement or diffusion, then spillovers violate the standard assumption of "no interference between units." Previous studies have generally ignored the possibility of interference between treatment and control hot spots, and focused instead on the spillovers into nearby non-hot spots. This is reasonable in small samples where hot spots are widely dispersed and the spillover regions do not overlap. But interference between units grows large as we scale up to hundreds of treated hot spots in a city. The same would be true of any intervention in a spatial or social network. This is a growing source of experimental work. We illustrate how to approach these challenges through the experiment design and randomization inference.

5.1 Design-based approach

We did not know the range of spatial spillovers, and so we pre-specified a flexible design that tested for spillovers in radii of 250m and 500m around treated streets.[26]

Our preferred approach partitions control segments into one of three experimental conditions according to their distance from the treated segment: <250m, 250–500m, and >500m. Figure 3 illustrates this partition. The hot spot segment at the center of the two radii was assigned to the intensive policing treatment. For simplicity Figure 3 ignores municipal services. Nearby hot spots are classified by their distance to the treated segment.

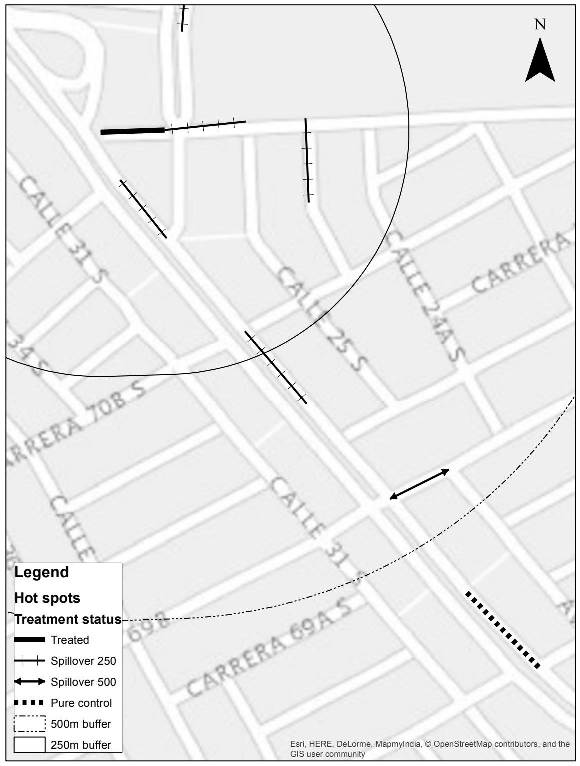
One virtue of this approach is that all treatment effects estimates are simply differences in the means of the experimental conditions in Table 2. We can also use this design to assess spillover effects on non-hot spots outside the experimental sample. We opt for regression-based estimates to control for possible confounders, as described below, but these preserve the spirit of the mean differences approach.

Our approach ignores the possibility of spillovers beyond 500m, as well as non-spatial spillovers. Some crime is undoubtedly displaced in non-Euclidean ways (e.g., to possibly distant hot spots where the benefits of crime are high and the risk of detection is low).[27]

[26]For details on all pre-specified aspects of the design see https://www.socialscienceregistry.org/trials/1156. There are many other ways to model spillovers, and we test robustness to a continuous rate of decay, as well as different radii. Previous literature on hot spots policing has focused mainly on catchment areas of about two blocks or 150m (Braga et al., 1999; Braga and Bond, 2008; Mazerolle et al., 2000; Taylor et al., 2011; Weisburd and Green, 1995). We felt 150m to be too conservative, however, and opted for 250m instead. We also specified a 500m option in case spillovers were unexpectedly large. Wider radii seemed implausible and would have eliminated the pure control category in a single city.

[27]Ferraz et al. (2016) find evidence of non-spatial spillovers in Rio de Janeiro's favela pacification. We

Figure 3: An example of assignment to the four treatment conditions



5.2 Randomization procedures

We used a two-stage randomization procedure to maximize the spread between hot spots assigned to each experimental condition. This ensured as many segments as possible had a high probability of assignment to the 250–500m and >500m conditions. We first blocked our sample by the 19 police stations, then randomized hot spots to intensive policing in two stages: first assigning quadrants to treatment or control, then assigning hot spots within treatment quadrants. We assigned no more than two hot spots per quadrant to intensive policing. This procedure assigned 756 hot spots to intensive policing and 1,163 to control.[28]

In March, we selected streets for municipal services. We sent enumerators to take five photographs and rate hot spots for the presence of disorder.[29] Of the 1,534 segments they were able to safely visit, 70% had at least one maintenance issue. We made these, plus

expect these non-spatial spillovers could lead us to overstate direct treatment effects and understate total spillovers.

[28]Within each station we took all quadrants with at least one hot spot and randomized quadrants to treatment with 0.6 probability. We then used complete randomization to assign hot spot segments to treatment within treatment quadrants.

[29]They looked for graffiti, garbage, and run-down buildings. A limitation is that we measured disorder after two months of policing treatment. We had no reason to expect the treatment to affect physical disorder, and there is no statistically significant difference between hot spot and non-hot spot streets.

Table 3: Distribution of treatment and spillover assignments across the experimental sample

		Municipal services assignment to:					
		Treatment	<250m	250m-500m	>500m	*Ineligible*	All
	Treatment	75	196	192	293	174	756
Intensive	<250m	74	281	185	165	162	705
policing	250m-500m	32	47	102	113	75	294
assignment	>500m	20	22	16	106	49	164
	All	201	546	495	677	460	1,919

Notes: "Ineligible" segments are those having no observed garbage or broken lights. For simplicity, we ignore whether ineligibles are <250m to hot spot policing or municipal services segments or not.

the 385 segments they could not visit safely, eligible for municipal services assignment. We blocked on police station and the previous intensive policing assignment, and assigned 201 hot spots (14% of eligible segments) to municipal services.[30]

Table 3 summarizes how the hot spot segments in our experimental sample are distributed across 20 treatment conditions and potential outcomes—4×5 conditions tied to the four conditions for each intervention (treatment, <250m, 250-500m, and pure control) plus the ineligible category of streets that we deemed were in no need of municipal services.[31]

Tests of randomization balance Random assignment produced the expected degree of balance along covariates. Table 1 reports the weighted means for a selection of baseline covariates, by experimental assignment, for hot spots and non-hot spots. For the most part, background attributes appear balanced across experimental conditions. There are some minor differences between treatment and control hot spots (for instance, treated hot spots are slightly less likely to be in industrial zones), but overall the imbalance is consistent with chance and is robust to alternative balance tests.[32]

[30] These 201 were the first "batch" to be treated. We also randomized a second batch of 214 hot spots for later treatment should the city decide to expand services. Two months into treatment of the first batch, however, our analysis of compliance records and visual inspection of hot spots suggested that continued municipal services were needed to maintain order in the first batch, and so the city did not give contractors the list of segments in the second batch. Thus the second batch remains in our control group.

[31] Technically there are 3×4 "ineligible" conditions, since streets that were diagnosed as having no need for municipal services could be <250m, 250–500m, or >500m from either treatment.

[32] To see whether covariate imbalance lies within the expected range, we test the null hypothesis that the covariates do not jointly predict experimental assignment. We use multinomial logistic regression with randomization inference to model the four-category experimental assignments for hot spots in the experimental sample (treatment, <250m, 250-500m and >500m), or the three-category assignments for streets in the non-experimental sample (<250m, 250-500m and >500m). To obtain exact p-values, we use randomization inference. Using simulated random assignments, we obtain a reference distribution of log-likelihood statistics under the null hypothesis; we then calculate the p-value by locating the actual log-likelihood value within this reference distribution. The p-value is non-significant, as expected, for both the experimental and non-experimental samples: $p = 0.681$ for hot spots and $p = 0.531$ for non-hot spots. We draw similar

5.3 Estimation

We estimate treatment and spillover effects within the experimental sample using the following weighted least squares regression:1

$$Y_{sqp} = \beta_1 P_{sqp} + \beta_2 M_{sqp} + \beta_3 (P \times M)_{sqp} + \lambda_1 S^P_{sqp} + \lambda_2 S^M_{sqp} + \lambda_3 (S^P \times S^M)_{sqp} + \gamma_p + \Theta X_{sqp} + \epsilon_{sqp} \quad (1)$$

where Y is the outcome in segment s, quadrant q and police station p; P is an indicator for assignment to intensive policing; M is an indicator for assignment to municipal services; S^P and S^M are indicators for the relevant spillover region (either <250m or <500m from treatment, or a vector of both indicators); γ is a vector of police station fixed effects (our randomization strata); and X is a vector of pre-specified baseline control variables.[33] Weights are the inverse probability weights (IPWs) of assignment to each experimental condition.

To calculate spillovers in non-hot spots we estimate:

$$Y_{sqp} = \lambda^N_1 S^P_{sqp} + \lambda^N_2 S^M_{sqp} + \lambda^N_3 (S^P \times S^M)_{sqp} + \gamma^N_p + \Theta^N X_{sqp} + \epsilon^N_{sqp} \quad (2)$$

using IPW for assignment to the conditions S^P and S^M. Thus, β_1 and β_2 estimate the marginal intent-to-treat (ITT) effects of each treatment alone and β_3 estimates the marginal effect of receiving both. A negative sign on β_3 implies increasing returns. The effect of receiving both interventions is the sum, $\beta_1 + \beta_2 + \beta_3$. Likewise, λ and λ^N estimate spillover effects of each treatment in each sample. To see the marginal effects of each treatment, we can perform the estimation under the constraints that $\beta_3 = 0$ and $\lambda_3 = 0$. These constraints are useful when we expect no interaction, such as the analysis of treatment compliance.[34]

Appendix E describes a model for estimating a continuous rate of decay in spillovers.

Inverse probability weighting Spillovers introduce spuriousness that can be corrected with IPWs. Hot spots close to other hot spots, such as those in the city center or other dense

conclusions from tests of treated vs control units >250m away and between control units <250m and >250 away.

[33]We selected these covariates by their ability to predict baseline crime levels. X also includes an indicator for segments ineligible for municipal services treatment by virtue of their baseline disorder.

[34]This estimation strategy represents a slight departure from the pre-analysis plan. The plan indicated that we would first and foremost focus on pairwise comparisons of each intervention separately, dropping from the regression any segments with a zero probability of assignment to any of the conditions. That approach generates similar results but, in retrospect, is problematic. Most importantly, a pairwise comparison of streets that did and did not receive intensive policing (ignoring municipal services treatment) would be biased since assignment to municipal services is slightly imbalanced across intensive policing experimental conditions (see Table 1). Hence we must estimate the effects of both interventions jointly. In addition, our original approach required us to drop an increasing number of segments from the regression, especially when estimating the interaction, rather than using the full sample. Equations (1) and (2) maintain the spirit of the original estimation approach but correct for these problems.

areas, will be assigned to the spillover condition in most randomizations. These streets may have unobservable characteristics that are associated with high levels of crime. This could mechanically lead us to conclude that spillovers increase crime. Controlling for baseline characteristics and crime histories reduces but does not eliminate the potential bias. With IPWs, outcomes for the segments assigned to any given condition are weighted by the inverse of the probability of assignment to that condition.[35] These weights, for instance, ensure that all segments have the same probability (after weighting) of being exposed to spillovers.

Procedure for determining the spillover condition To determine the relevant spillover radii for conditions S^P and S^M, we pre-specified a procedure: if there is no evidence of statistically significant spillovers into the 250–500 m region using a $p < .1$ threshold, then S^P and S^M will indicate segments in the <250m spillover region only, otherwise they will indicate segments <500m of treated hot spots. If there are no statistically significant spillovers in the 250m radius nor the 250-500m radius, then our primary estimates would ignore the classification of control streets into various spillover conditions and estimate the β coefficients alone. In retrospect, our pre-specified rule for determining the spillover range was too permissive. First, it was based on spillovers in the experimental sample rather than the much larger non-experimental sample. Second, this rule could lead us to ignore quantitatively large but imprecisely estimated spillovers. In principle it could lead us to ignore spillovers large enough to offset any direct benefits of crime reductions in treatment hot spots. Thus we will also show results accounting for spillovers into non-hot spots.

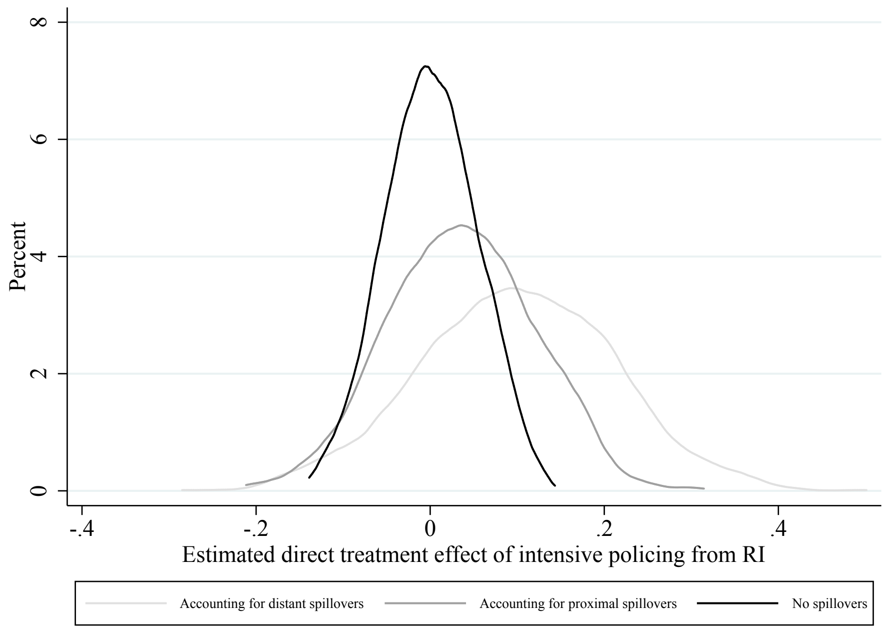
Overall, the above approach is similar to the approach that previous studies have used to estimate spillovers into a nearby catchment area. Our advantages include: we can estimate spillovers flexibly over various radii; we can account for overlapping catchment areas of both the hot spots policing intervention and the municipal services intervention; and we can estimate exact p-values.

5.4 Why randomization inference?

Randomization inference (RI) gives precise p-values based on the empirical distribution of all estimated treatment effects that could arise under our design and data under the null hypothesis of no effect for any unit. RI reassigns treatment randomly thousands of times, each time estimating the treatment effect that could have arisen by chance from that comparison.

[35] Each segment's probability of exposure to <250m or 250-500m spillovers can be estimated with high precision by simulating the randomization procedure a large number of times. Such IPWs have a long history in survey sampling and have become common in the analysis of randomized trials with varying probabilities of assignment (Horvitz and Thompson, 1952; Gerber and Green, 2012). Appendix C describes and maps IPWs in our sample.

Figure 4: The empirical distribution of estimated treatment effects on insecurity under different spillover scenarios



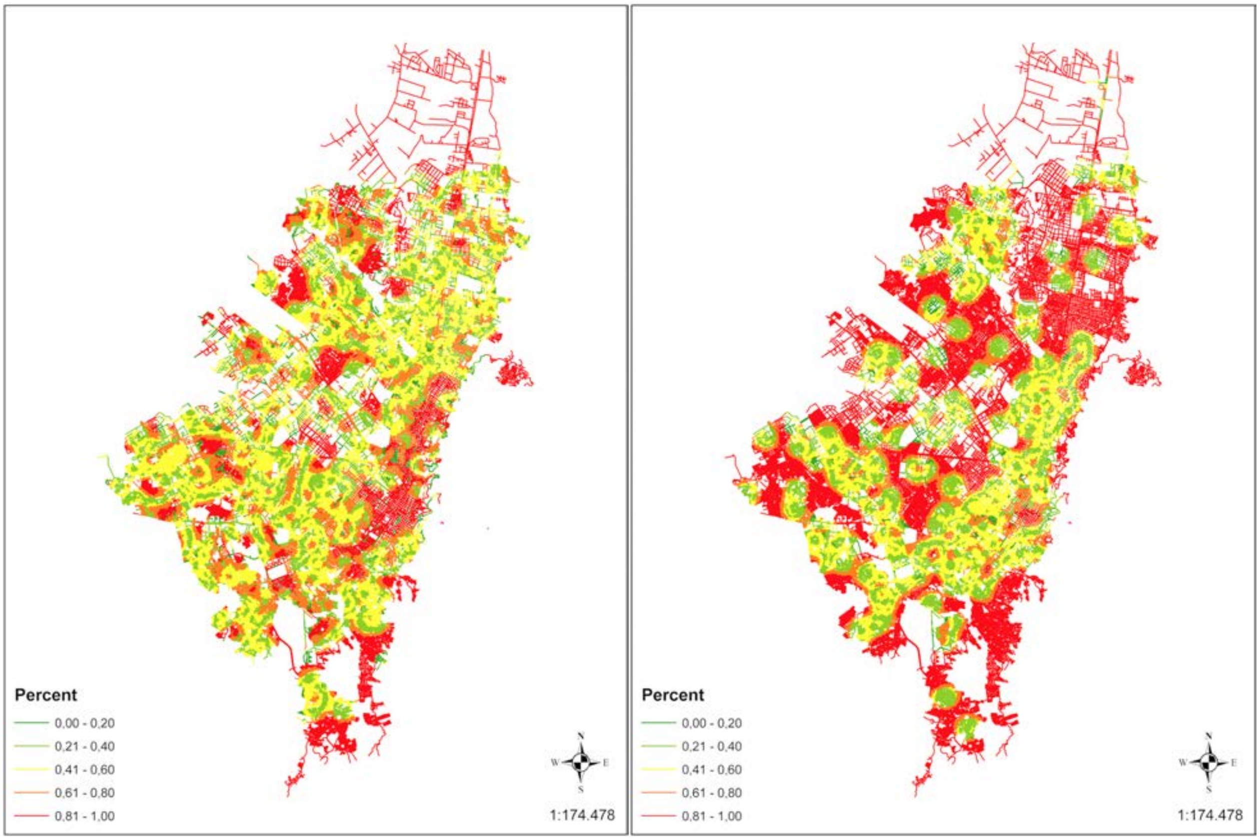
Notes: The figure displays the empirical distribution of treatment effects on the insecurity index for intensive policing. We simulate the randomization procedure 1,000 times and estimate treatment effects for each randomization using post-treatment data under the sharp null of no treatment effect for any unit. The figures show distributions for three cases of equation (1): the simple treatment-control comparison with no spillovers (i.e. $S_s^P = S_s^M = 0$ for all s); the case where S^P and S^M indicate proximal spillovers within 250m; and the case where S^P and S^M indicate the larger spillover area within 500m.

Figure 4 displays the empirical distributions of treatment effects for intensive policing under three cases of equation (1): the simple no-spillovers case (i.e. $S_s^P = S_s^M = 0$ for all s); the case where S^P and S^M indicate spillovers within 250m only; and the case where S^P and S^M indicate spillovers within 500m.

Most importantly, the distribution widens when accounting for spillovers. The no-spillovers case has the narrowest distribution, and the p-values associated with each treatment effect are nearly identical to the p-values obtained from conventional WLS standard errors clustered by randomization strata. The distribution widens as we account for wider spillover regions. We are more likely to get large treatment effects by chance. Thus the standard errors estimated by the WLS regressions (1) and (2) will be too small.

This widening of the sampling distributions follows from two facts. One is that we are losing data as we pare off rings of spillovers. The second is that the control region becomes more separate from the region where the treatment hot spots are located. Hot spots that are close to other hot spots are assigned to the spillover condition in most randomizations, creating patterns of "fuzzy clustering" (Abadie et al., 2016). These clusters are difficult to model because they have to do with distance from other hot spots rather than an observed

Figure 5: Fuzzy clustering in the presence of spillovers



Notes: The figure displays the proportion of segments within 500m assigned to the same treatment condition for intensive policing (left) and municipal services (right).

characteristic such as a quadrant.

We can see the fuzzy clustering in Figure 5, which illustrates for each segment the proportion of segments within 500m that are assigned to the same experimental condition, including the spillover conditions. For instance, for intensive policing, most segments in the dense city center (the middle right of the map) have neighbors with the same high probability of assignment to the <250m spillover condition. For municipal services, there are large swathes of the city with a high probability of assignment to the control condition, forming a cluster that does not conform to administrative boundaries. The figures imply that, instead of having thousands of independent segments, we actually have dozens of clusters. But there is no geographic marker for them.

Finally, the simulations in Figure 4 show that the distribution of simulated treatment effects with spillovers are not centered at zero. Equations (1) and (2) can lead to a small level of bias in estimated coefficients, even when using IPWs. Clustered assignment introduces bias when there are clusters of unequal size, and when cluster size is correlated with potential outcomes. When we ignore spillovers, we stipulate that there is no such clustering, which is why that distribution is centered at zero. When we allow for spillovers, we confront the fact that our exposure to spillovers is clustered. The bias disappears as the number of clusters increases (and indeed it is negligible when we estimate non-experimental spillovers). Unfortunately, the spillover effects we estimate will often be subtle, and so the bias is fairly large in comparison to some of the direct average treatment effects.

What RI allows us to do is to assign a p-value for a given treatment effect by observing where that treatment effect falls in the distribution of all possible effects in 10,000 randomizations. We use these RI p-values in place of the conventional standard errors and p-values whenever we estimate treatment effects in the presence of spillovers. The simulations used in the RI procedure provide an estimate of the bias. All of our tables report bias-corrected treatment effects. Appendix C reports the specific biases estimated. We will also report estimates of treatment effects without weights and randomization inference in the Appendix.

6 Results

6.1 Program implementation and compliance

The police patrols and municipal services complied with instructions and treatment assignment.[36] Police did so for the full eight months, while municipal services agencies likely complied for a shorter period. Table 4 reports the effects of assignment to intensive policing or municipal services on various first-stage outcomes. We estimate equation 1 ignoring interactions between the two treatments (we have no reason to expect one treatment to affect compliance with another). For simplicity we compare treatment segments to all control segments, ignoring spillovers. Accounting for spillovers yields similar conclusions (not shown).

Intensive policing Calculating the time spent on street segments is difficult because of periodically malfunctioning units or outages. We estimate control streets received 92 minutes of patrolling time per day, on average. Treated streets received an extra 77 minutes, a 84% increase.[37] Streets outside the experimental sample received an average of 33 minutes of patrolling time per day. Without pre-treatment data on patrol times it is impossible to say whether the increase in patrol time on treatment hot spots came at the expense of control hot spots. What we can say is that the 77 minute rise on two segments means roughly a

[36]This and all other results went through a pre-publication re-analysis process by the J-PAL Research Team.

[37]If we account for potential spillovers, streets within 250m of treated hot spots received about 7 more minutes of patrol time compared to more distant segments, $p = 0.18$ (not shown)

Table 4: "First-stage" effects of treatment on measures of compliance and effectiveness

Dependent variable	Control mean	ITT and standard error of assignment to: Intensive policing		Municipal services	
	(1)	(2)	(3)	(4)	(5)
A. Intensive policing measures:					
Proportion of respondents who say police presence increased in last 6 mo.	0.129	0.076	[.011]***	0.017	[.013]
Daily average patrolling time, excluding quadrant-days without data	92.001	76.571	[4.424]***	-3.333	[4.371]
# of arrests	0.333	-0.053	[.082]	0.026	[.102]
# of drug seizure cases	0.041	-0.002	[.020]	0.029	[.024]
# of gun seizure cases	0.009	0.006	[.008]	0.007	[.013]
# of recovered car cases	0.003	0.000	[.001]	-0.003	[.001]*
# of recovered motorbike cases	0.006	-0.028	[.019]	0.032	[.027]
B. Municipal services implementation measures			0.006		0.016
Proportion of respondents who say municipal presence increased in last 6 mo.	0.144	0.006	[.010]	0.016	[.012]
City determined segment is eligible for lights intervention	0.349	-0.007	[.048]	-0.139	[.048]***
Received lights intervention	0.000	-0.010	[.020]	0.199	[.026]***
City determined segment is eligible for garbage intervention	0.000	0.011	[.025]	0.627	[.032]***
Received garbage intervention	0.000	0.015	[.026]	0.382	[.033]***
June 2016 enumerator assessment of street conditions:			-0.018		0.078
Graffiti on segment	0.749	-0.018	[.050]	0.078	[.043]*
Garbage on segment	0.251	0.071	[.061]	0.015	[.049]
Visibly broken street light on block	0.000	0.012	[.012]	0.008	[.008]
December 2016 enumerator assessment of street conditions:			0.019		0.059
Graffiti on segment	0.624	0.019	[.053]	0.059	[.047]
Garbage on segment	0.245	0.021	[.051]	0.002	[.043]
Visibly broken street light on block	0.029	0.022	[.016]	-0.015	[.017]

Notes: This table reports intent to treat (ITT) estimates of the effects of the two interventions, via a WLS regression of each outcome on treatment indicators, police station (block) fixed effects, and baseline covariates (see equation 1, where we have constrained the coefficient on the interaction term to be zero and ignored spillovers). The regression ignores spillover effects. Standard errors are clustered using the following rules: (i) for all treated segments except with cluster size 2, each segment is a cluster; (ii) for all other untreated segments, each segment gets its own cluster identifier; (iii) for entirely untreated quadrants, they form a cluster; and (iv) for quadrants with exactly 2 units assigned to treatment, those units form a cluster. The proportion of people reporting increased state presence comes from our citizen survey, the enumerator assessments were collected by the research team, and the remainder of the outcomes come from police administrative data. * significant at the 10 percent, ** significant at the 5 percent, *** significant at the 1 percent.

Table 5: Municipal services eligibility and compliance

		City's lighting assessment			% of eligible streets
		Lights eligible	Lights ineligible	All	receiving lighting service
City's	Eligible for garbage	21	102	123	41 (89.1%)
cleanliness	Ineligible for garbage	26	52	78	
assessment	All	47	154	201	
% of eligible streets receiving clean-up		74 (60.2%)			

Notes: The table summarizes compliance on the municipal services intervention for 201 streets assigned to treatment as reported by the corresponding agencies within the Mayor's office.

minute less time on each of the 130 other segments in the quadrant. Some citizens noticed an increase in patrols in the previous 6 months. On control segments 13% reported an increase, compared to 21% on treatment segments.

We do not see any effect of increased policing on arrests or police actions such as drug seizures, suggesting any effect of the policing may be through deterrence rather than incapacitation (Chalfin and McCrary, 2017).

Municipal services Table 5 summarizes compliance. After assigning 201 segments to municipal services, city agencies diagnosed each one in March. They identified 123 segments needing clean-up services, and 47 needing lighting improvements. They performed the services June through August. Tree pruning and graffiti cleaning were one-time treatments; rubbish collection was expected to be semi-regular. Based on city data, 74 of the 123 streets (60%) were cleaned up, and in 41 of the 47 streets (87%) they repaired broken lights and replaced poor lights with better ones. No graffiti was cleaned-up.

The impacts were not obvious. 14.4% of survey respondents on control segments noticed an improvement in service delivery in the past six months, and this was only 1.9 percentage points greater in treatment streets (not statistically significant, see Table 4). We also visited segments in daytime in June and December 2016 to photograph and rate the streets. The before and after photos generally display relatively tidy streets and before-after differences are imperceptible. It is possible that lights repairs were more evident, but it was unsafe to visit hot spots at night. We see no effect of treatment in Table 4. One possibility is that the extensive margin is the wrong margin to evaluate, and another is that the disorder in cleaned up segments could have re-accumulated over days or weeks.

6.2 Program impacts on officially reported crime

We begin by analyzing impacts using administrative crime data from all streets in the city. Table 6 reports results from estimating the direct treatment (β), experimental spillover (λ), and non-experimental spillover (λ^t) coefficients from equations (1) and (2), with and without the interaction terms between intensive policing and municipal services.[38] Following our pre-specified rule, the table reports spillovers within 250m only. We do not see statistically significant spillovers in the 250–500m region (Appendix E). The Appendix also reports an alternative, more general test demonstrating spillovers within 250m but not in the 250–500m region.

Table 6 also calculates the total number of deterred crimes, as the product of the estimated coefficients and the number of treatment and spillover segments in the city. We omit the 57,695 streets with zero probability of assignment to the spillover condition. There are 51,390 non-hot spots and 705 control hot spots for the policing intervention and 20,740 non-hot spots and 546 control hot spots for municipal services. Thus even small estimated spillovers can have a large effect on the total crime estimates. Since our coefficients are fairly uncertain, we must take aggregate impacts with caution.

Our best guess for the overall impact on crime is that the interventions directly deters a relatively modest amount of crime, and that some or all of this crime is displaced to neighboring streets. However, in our main specification, crime displacement is concentrated in property crime. Violent crimes may not be displaced so easily.

Direct treatment effects Starting with columns 1–4 of Table 6 (no interaction), both intensive policing and municipal services reduce officially reported crimes on average, although these coefficients are not statistically significant. Control segments report an average of 0.743 crimes over the intervention period (column 1 in Table 8). Thus the coefficient on intensive policing of -0.094 represents a 12.6% improvement. The municipal services coefficient is about two-thirds as large. In total, these estimates suggest that the reallocation of police and municipal services deterred 86 crimes in targeted streets over the intervention period (not statistically significant).

Turning to columns 5–8, we see larger and most statistically significant impacts of state presence in the segments that were assigned to both interventions. The coefficients on policing and municipal services are positive but imprecise. We see no evidence that either intervention on its own reduced crime. The coefficient on the interaction is -0.437, however,

[38] We pre-specified a one-tailed test since we had strong priors about the direction of the effect. But significance levels in the table reflect a two-tailed test to be conservative and consistent throughout. As noted above (footnote 34) we also pre-specified a pairwise analysis for treatment effects. While this proved to be an erroneous choice, Appendix D reports those pre-specified pairwise results.

Table 6: Estimated aggregate impacts of the interventions, accounting for spillovers within <250m

	Dependent variable: # of crimes reported to police on segment (administrative data)							
	No interaction between treatments				Interaction between treatments			
Impacts of treatment	Coeff. (1)	RI p-value (2)	# segments (3)	Estimated total impact = (1) × (3) (4)	Coeff. (5)	RI p-value (6)	# segments (7)	Estimated total impact = (5) × (7) (8)
A. Direct treatment effect								
Intensive policing	-0.094	*0.512*	756	-70.7	0.009	*0.817*	756	7.2
Municipal services	-0.076	*0.782*	201	-15.2	0.089	*0.367*	201	17.9
Both					-0.437	***0.043***	75	-32.8
Subtotal				-85.9				-7.8
B. Spillover, experimental sample								
Intensive policing	0.061	*0.595*	705	42.7	0.143	*0.315*	705	100.6
Municipal services	0.176	***0.056***	546	96.3	0.255	***0.025***	546	139.2
Both					-0.272	*0.196*	281	-76.5
Subtotal				138.9				163.3
C. Spillover, non-experimental sample								
Intensive policing	0.016	*0.113*	51390	840.7	0.013	*0.219*	51390	654.1
Municipal services	-0.003	*0.416*	20740	-55.5	-0.006	*0.500*	20740	-120.8
Both					0.006	*0.968*	15491	95.7
Subtotal				785.2				629.0
Net increase (decrease) in crime				838.2				784.5
			95% CI	(−813, 2131)			*95% CI*	(−1063, 2268)
			90% CI	(-492, 1919)			*90% CI*	(-735, 2033)

Notes: Columns 1–4 refer to the non-interacted results (equation 1 under the constraint that $\beta_3 = 0$ and $\lambda_3 = 0$) while columns 5–8 refer to the interacted results (equation 1 with no constraints). Columns 1 and 5 display the bias-adjusted treatment effect while columns 2 and 6 display RI p-values. Columns 3 and 7 display the number of units in each group. Columns 4 and 8 display the product of the bias-adjusted treatment effect and the number of units in each group. The confidence interval on the bottom of the table is constructed using randomization inference. First we create a fake schedule of potential outcomes for each observation by adding or subtracting RI-adjusted treatment or spillover effects. This process gives us a potential outcome for each unit depending on its treatment assignment. Second, we simulate a randomization and take the potential outcome associated with the treatment assignment of the new randomization. Third we estimate treatment and spillover effects using this new outcome and apply the RI bias adjustment from our main set of results. Fourth, we multiply these bias-adjusted treatment effects by the number of segments in each group, and sum across both the experimental and non-experimental samples to get the aggregate effect. We repeat steps two through four 1,000 times to get the distribution of the test statistic, which is roughly centered on the actual number of deterred crimes. The 2.5 and 97.5 percentiles of this distribution give us the 95% confidence interval. $p < .1$ in bold.

with a RI p-value of 0.043. The sum of the three coefficients is -0.339 with a p-value of 0.109 (see column 5 in Table 8). This sum corresponds to a 45.6% decrease in reported crimes on the 75 streets that received both interventions. The fact that the coefficient on the interaction is large, negative, and statistically significant implies that there may be increasing returns to security investments, at least over this range of variation. Of course, given that the sum of effects is weakly statistically significant, we cannot say with confidence that both interventions reduced crime on these 75 streets. Moreover, the aggregate direct effect of the program looks even smaller when we account for the interaction. According to these estimates, our best guess is that only 8 crimes were deterred directly by both interventions—about 1 per month of the policing and municipal services interventions.

Spillover effects Meanwhile, the spillover coefficients suggest that any crime deterred is more than made up for by a rise in crime in streets within 250m. For intensive policing, all four spillover coefficients are positive. The spillover effects in the experimental sample are imprecise, but given the large number of nearby non-hot spots, the spillovers in the non-experimental sample suggest a positive effect, albeit one that falls short of conventional levels of significance (even jointly). There are a sufficiently large number of non-hot spot segments that these small coefficients add up to high levels of crime—841 crimes in aggregate when we do not allow for the interaction and 654 when we do. In contrast, we see no evidence that municipal services pushed crime around the corner. The coefficients on spillovers in the non-experimental sample are actually negative, though they are imprecise. In aggregate, however, this estimate adds up to between 56 and 121 crimes deterred in nearby streets, depending on the specification.

Aggregate effects We use these estimates to guess the aggregate effect on crime. It is unlikely that reallocating police and municipal services reduced total crime in the city. On the contrary, the estimates suggest crimes increased by about 800 in both specifications (2% relative to the total number of reported crimes). This must be taken with caution, however, for two reasons. First, neither aggregate effect is statistically significant at even the 10% level. Second, this estimate would not capture general equilibrium effects if they exist (e.g. if the intervention is disrupting city-wide criminal networks). These estimates suggest that we can rule out the possibility that crime decreased in the city by even a modest amount.

Robustness to choice of spillover region The choice of spillover region is somewhat sensitive to the particular test. Appendix E reports our preferred test, a joint test of differences in means across experimental conditions. This suggests we should focus on spillovers within 250m. In retrospect, our pre-specified test for spillovers should have accounted for large estimates where the coefficient is slightly above p=0.1, interactions between treatments, as well as baseline covariates. In general, however, the alternative methods lead to similar conclusions: small direct effects of the interventions on crime, and evidence that policing displaces crime to nearby streets, with generally large confidence intervals.

As the spillover region shrinks, the magnitudes of the direct effects of policing and municipal services remain similar but precision improves. In the extreme case, where there are no spillovers, the direct effects are statistically significant but suggest that in aggregate relatively few crimes are deterred.

We consider an exponential rate of decay rather than our fixed radii (see Appendix E.3). There are two differences in results. One is that municipal services appears to have positive spillover effects. The impact of municipal services on nearby streets becomes significant at almost the 10% level, and at almost the 5% level for violent crimes. Second, the negative effects of intensive policing on nearby streets is no longer confined to property crime. Now violent crime appears to be pushed around the corner.

Heterogeneity by type of crime Police prioritize violent crimes over property crimes.[39] Table 7 disaggregates the impacts on total crime into violent and property crimes. Our best guess is that aggregate violent crimes fell by 135 to 374 crimes in total (1% to 3% relative to the total number of violent crimes), depending on whether we use the interaction or not, although neither estimate is statistically significant. Property crimes rose by 1,014 to 1,205 in aggregate (4% to 5% relative to the total number of property crimes), however, and these estimates are statistically significant at the 10% level when we include the interaction. The two most socially costly crimes, homicides and sexual assaults, fall by 65 to 97 crimes (5% to 8% relative to the total number of homicides and sexual assaults). This difference in property and violent crimes is statistically significant. See Appendix F.1 for detailed results.

We take the different results for violent and property crimes with caution, however, since the aggregate effects change once we introduce minor changes in specification. We estimate an alternative version of equation (1), with different dummies for streets located within 250m and between 250-500m from treatment hot spots. See Appendix F.2. In this case, we now observe crime displacement also for violent crime.

Heterogeneity by initial level of crime We pre-specified one major form of heterogeneity analysis, by baseline levels of crime. Broadly, we observe what we predicted: that improvements in insecurity are greater in the higher-crime streets. Figure 6 reports the

[39]More specifically: murder, rape and assaults over other crimes such as burglary or car theft.

Table 7: Aggregate impacts on crimes by type (mean and confidence intervals)

	without interaction			*with interaction*		
	Effect	95% CI	90% CI	Effect	95% CI	90% CI
	(1)	(2)	(3)	(4)	(5)	(6)
All crime	838.2	(–813, 2131)	(-492, 1919)	784.5	(–1063, 2268)	(-735, 2033)
Property crime	1014.4	(-195, 2075)	(-44, 1903)	1205.1	(-340, 2385)	(23, 2239)
Violent crime	-135.0	(-853, 389)	(-747, 281)	-374.1	(-1134, 213)	(-1011, 75)
Homicides and sexual assaults only	-65.3	(-178, 55)	(-162, 41)	-97.1	(-236, 32)	(-210, 16)
Property – violent crime	1149.3			1579.1		
p-value	0.068			0.018		

Notes: This table presents the aggregate effect calculation for various crime subgroups assuming spillovers within 250m. Calculations are based on the aggregate effect and confidence interval described in Table 6

results of estimating treatment effects on the n% highest crime hot spots. The treatment effect is fairly constant up until the point we reach the street segments in the 70^{th} percentile and above, when the impact of receiving both interventions climbs first to 0.5 standard deviations and then to about 0.75 standard deviations. The effect is imprecise, as the sample size drops dramatically. These results are consistent with increasing returns to treating the least secure hot spots.[40]

6.3 Program impacts on insecurity

Table 8 reports impacts on our main security measures: the perceived risk index, based on surveys; and the index of crime, which averages survey- and officially-reported crime. Treatment effects can be interpreted as average standard deviation changes in the outcome. The table also reports treatment effects on components of the crime index. Our focus is on the two pre-specified indexes, but we also report results for an equally-weighted average of both. Table estimates equation (1), and reports direct treatment effects and spillover effects on hot spots within 250m.

The survey data tell a similar story as police data. We see the largest and most statistically significant impacts of state presence in the segments that received both interventions. Those 75 segments reported a 0.327 standard deviation decrease in overall insecurity, significant at the 10% level (column 5). The coefficients on perceived risk and crime indexes are similar, though only the perceived risk index is statistically significant alone.

Alone, the interventions are associated with improvements in security, but none of the estimates are individually significant. Nonetheless the coefficients all point in the direction of

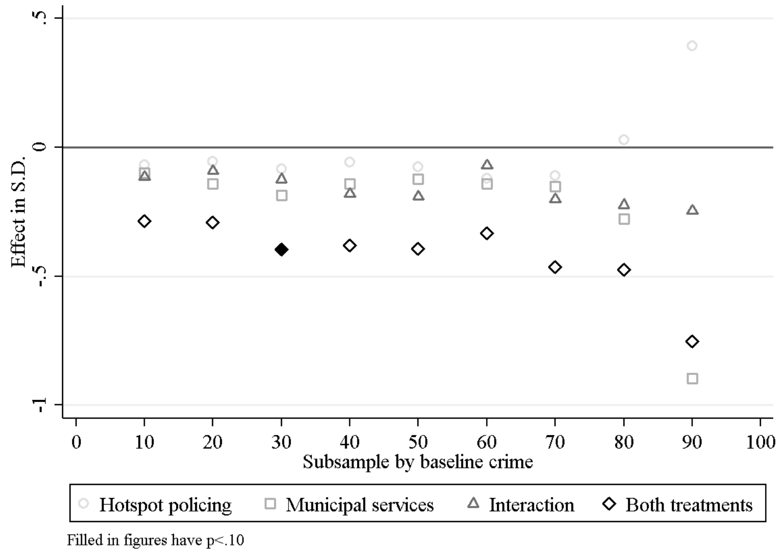
[40] As we show in Appendix F, the cumulative effect of both interventions appears within 8 to 12 weeks of the intervention, and grows over time.

Table 8: Program impacts on security in the experimental sample, accounting for spillovers within 250m, with p-values from randomization inference (N=1,916)

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (2), (3), and (4)	Impact of spillovers <250m: Any intensive policing	Any municipal services	Both interventions	Sum of (6), (7), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Insecurity index, z-score (+ more insecure)	-0.003	-0.106	-0.100	-0.121	-0.327	0.045	0.150	-0.217	-0.022
		0.391	*0.536*	*0.447*	***0.095***	*0.322*	***0.020***	***0.039***	*0.577*
Perceived risk index, z-score (+ riskier)	0.049	-0.122	-0.086	-0.084	-0.292	0.002	0.083	-0.160	-0.075
		0.259	*0.494*	*0.644*	***0.094***	*0.511*	*0.129*	***0.085***	*0.808*
Crime index, z-score (+ more crime)	-0.054	-0.054	-0.080	-0.118	-0.252	0.073	0.166	-0.200	0.039
		0.701	*0.659*	*0.412*	*0.196*	*0.231*	***0.010***	***0.059***	*0.361*
Perceived & actual incidence of crime, z-score	0.059	-0.081	-0.158	0.066	-0.173	0.027	0.099	-0.137	-0.011
		0.514	*0.153*	*0.423*	*0.507*	*0.418*	***0.092***	*0.171*	*0.578*
# crimes reported to police on street segment	0.743	0.009	0.089	-0.437	-0.339	0.143	0.255	-0.272	0.125
		0.817	*0.367*	***0.043***	*0.109*	*0.315*	***0.025***	*0.196*	*0.289*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports intent to treat (ITT) estimates of equation 1, estimating the direct effects of the two interventions (Columns 2 to 4) and the spillover effects (Columns 6 to 8) via a WLS regression of each outcome on treatment indicators, spillover indicators, police station (block) fixed effects, and baseline covariates. Columns 5 and 9 report the sum of the three preceding coefficients. The measures of perceived risk, perceived incidence of crime, and proportion reporting crime come from our citizen survey, and the # of crimes reported to the police come from police administrative data.

Figure 6: Heterogeneity of security impacts by pre-treatment administrative crime levels



Notes: We estimate equation 1 nine times, each time interacting each treatment indicator with an indicator for whether a segment is below the nth percentile of baseline crime levels among our experimental sample of hot spots, for n = 10, 20, ... , 90. The coefficients on the treatment indicators indicate the effect on the higher crime segments above that percentile (hence the right side of the scale is highest crime). The figure graphs these coefficients.

better security: intensive policing alone reduces perceived risk by 0.12 standard deviations, crime by 0.06, and overall insecurity by 0.11; while municipal services alone reduces perceived risk by 0.09 standard deviations, crime by 0.08, and overall insecurity by 0.10. The coefficient on the interaction term (column 4) is statistically significant for officially reported crimes only. We take this result as being suggestive of increasing returns to state presence.

Spillovers There is also evidence of crime displacing to control hot spots in columns 6 to 9 of Table 8. Intensive policing alone and municipal services alone are associated with increases in crimes on nearby hot spots of 0 to 0.26 standard deviations. Only the municipal services impacts are statistically significant, with a 0.15 standard deviation increase in insecurity. The interaction terms are generally negative (see column 8) and generally statistically significant, such that there is generally no evidence of spillovers onto hot spots near to hot spots that received both intensive policing and municipal services.[41]

Robustness What would we have found if we ignored different probabilities of treatment and the unusual patterns of clustering? In Appendix F we estimate "naïve" treatment

[41] We also have survey data on 399 non-experimental street segments, and Appendix E estimates these non-experimental spillovers within 250m. This sample is generally too small to estimate non-experimental spillovers precisely, but the patterns are generally consistent with what we see in the large-sample dataset on reported crimes, in particular the positive coefficients on intensive policing are positive.

effects ignoring IPWs and randomization inference. Direct treatment effects are slightly smaller than our main results, but the patterns remain similar. In contrast, naïve spillover effects are larger and highly statistically significant. Hence failing to account for interference between units and clustering of treatment conditions would have led us to severely exaggerate the degree to which policing pushes crime elsewhere. We also report results using different spillover regions in Appendix E.

Underreporting of official crime is not correlated with treatment These survey data also provide an opportunity to test whether people were more likely to report crimes to the police on treated segments. If so this would call into question the analysis of administrative police data in the last section. We see no difference in crime reporting on treated streets. The survey asked respondents their likelihood of reporting a future crime to the police, on a scale of 0 to 3. The average response in control segments was 2.0, with a treatment effect [standard error] of 0.016 [.029] from policing and 0.035 [.032] from municipal services. This suggests that administrative data are suitable for outcome assessment even while the treatment is being delivered.

Disentangling municipal services Our qualitative work and compliance data hinted that the lighting intervention may have been more compliant, effective, and persistent than the street clean-up. But the data do not support this conclusion, however. Both services appear to have been important. First, we see no evidence that municipal services treatment effects were concentrated in the hot spots diagnosed as needing improved lights. Second, we don't see larger treatment effects at nighttime. See Appendix F for this analysis.

6.4 Program impacts on state trust and legitimacy

We pre-specified three secondary outcomes capturing impacts on trust in and legitimacy of the state. First, an *opinion of police index* averaging 4 attitudes towards police: trust; quality of work; overall satisfaction; and likelihood they would give information to police. Second, an *opinion of mayor index* that asks the same 4 questions for city government. Third, a crime reporting measure that captures the likelihood that people reported a crime to the police. This helps us understand whether administrative crime reporting changes with treatment, but is also a measure of collaboration and hence legitimacy.[42]

Overall, we see little evidence that the interventions increased trust in or legitimacy of the state. Table 9 reports ITT effects using equation 1. We see an unexpected pattern: intensive

[42] In the state building and especially the counter insurgency literatures such civilian information, tips, and collaboration are among the chief indicators of state legitimacy.

Table 9: Impacts on state legitimacy allowing spillovers within 250m, with RI p-values

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)	Impact of spillovers <250m: Any intensive policing	Any municipal services	Both interventions	Sum of (6), (8), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Opinion of police, z-score (+ better)	0.024	0.143	0.210	-0.308	0.045	-0.024	0.043	0.123	0.141
		0.150	*0.107*	***0.017***	*0.867*	*0.590*	*0.817*	*0.338*	*0.797*
Opinion of mayor, z-score (+ better)	-0.014	0.001	0.179	-0.414	-0.234	-0.024	0.068	-0.025	0.020
		0.912	***0.078***	***0.003***	***0.008***	*0.523*	*0.982*	*0.919*	*0.668*
Likeliness to report crime (0-3, + higher)	2.046	0.004	0.021	0.035	0.060	-0.007	0.007	0.026	0.026
		0.921	*0.800*	*0.522*	*0.385*	*0.688*	*0.991*	*0.638*	*0.837*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports intent to treat (ITT) estimates of equation 1, estimating the direct effects of the two interventions (Columns 2 to 4) and the spillover effects (Columns 6 to 8) via a WLS regression of each outcome on treatment indicators, spillover indicators, police station (block) fixed effects, and baseline covariates. Columns 5 and 9 report the sum of the three preceding coefficients. The three measures come from our citizen survey.

policing and municipal services alone are associated with increases in the opinion of police and Mayor, but this is effectively cancelled out when both treatments are received. This pattern is statistically significant when we ignore spillovers, but less robust when accounting for spillovers. This heterogeneity across arms is hard to interpret and could reflect noise. In analysis ignoring any interactions (not shown), intensive policing and municipal services are associated with little change in opinions of police, and a slightly negative effect on Mayoral opinion—a 0.13 standard deviation fall, significant at the 10% level.

7 Discussion and conclusions

Not surprisingly, we find that direct and targeted state presence deters crime and violence. We also find some evidence of increasing returns to state presence. What was surprising to us was the small number of total crimes deterred. Also important is the divergent patterns of spillovers. We see evidence that intensive policing pushed property crime around the corner. In our main specification, however, state presence seems to have reduced the most serious violent crimes: murder and rape. In spite of our large sample, confidence intervals are wide, especially on the aggregate effects. The reduction in violent crimes is also sensitive to specification. Thus our study is a good example of a policy evaluation where the implications hinge on how to interpret estimates and significance levels under uncertainty.

Cost-benefit considerations Cost-effectiveness in this case is in the eye of the beholder. The city sees the interventions as having little or no marginal cost, since they simply reallocated existing resources from some streets to others without raising their budgets or personnel. If so, then the main question is whether a high likelihood of reducing roughly 100 murders and rapes (8% relative to the total number of cases) is worth a rise in property crime. This is a trade off that many police chiefs and mayors might reasonably make.[43]

On the other hand, reallocating street-level bureaucrats had real costs. There was a logistical cost of coordinating patrols, especially management time. It also made police patrols spend more time in unpleasant places. Officers told us they disliked the loss of autonomy and flexibility. There are also opportunity costs to consider. Intensive policing was a major reform, and like any bureaucracy, the police can only undertake so many reforms in a year. The Mayor's office used scarce social and political capital to implement it. We believe one should measure this reform against the others it supplanted.

[43] Indeed, we run the aggregate effects reported in Table 6 using a weighted crime index as outcome instead of the simple sum (not reported). The weights are the average prison sentence in the Colombian penal code for each crime. For instance, the weight for one homicide is about 13 times that for a shoplifting case. In such a case, the aggregate effects are negative for the interacted version, although imprecise.

How do our results line up with the U.S. evidence? This experiment provides some of the first experimental evidence on place-based crime interventions outside the U.S.[44] At first glance, it might seem that the displacement of total crime to nearby streets runs against the U.S. literature. We have to compare with caution as Bogotá and the U.S. are different contexts. Policing interventions also take different forms, and vary in terms of intensity, concentration, crimes targeted, duration, and quality of approach. That said, on close inspection, our results are not so different.[45] The previous literature has not ruled out positive or negative spillovers in a definitive way. These studies split on whether they observe displacement of crime or diffusion of benefits on average. Moreover, most prior studies' sample sizes are so small that the confidence intervals on spillovers include sizable displacement effects.[46] Perhaps the biggest lesson for place-based crime studies is that small sample sizes will simply not help answer the crucial question of spillovers.

Methodological lessons We believe that what matters most about this Bogotá result is not whether it generalizes to the U.S. or not, or runs against the literature, but the methodological lesson for future policy experiments in dense networks of streets or people. When small spillovers matter, anything that could bias spillover effects or make them less precise matter a great deal. This points to the importance of eliminating these biases and having accurate, efficient estimates. Failure to account for the biases arising from spillover estimation will have profound effects on our conclusions, whether it is the bias correction through IPW and re-centering, or randomization inference for calculating exact p-values.

Randomization inference has yet to gain currency in randomized trials, in part because most times RI provides more or less the same conclusion as the usual clustered standard errors. A textbook case for randomization inference, however, is design-based estimation of spillovers where units have widely different probabilities of assignment to different experimental conditions. This problem extends to any other situation in which the structure of the clustering of experimental units in a given treatment condition is difficult to model, which is

[44] Two ongoing projects in Latin America and the Caribbean are Collazos et al. (2017) in Medellín, Colombia and Sherman et al. (2014) in Trinidad and Tobago. Compared to the Medellín study we find generally different results. We observe direct treatment effects on both property and violent crimes, while they only find evidence of a decrease on car thefts. We observe displacement mainly on property crimes, and they find a decrease in car thefts in places nearby targeted hot spots. The context is radically different, regarding both criminal behavior and implementation capability, and we believe this could be driving the differences. For instance, Medellín has about 60% more police than Bogotá in relative terms.

[45] Most previous studies use only post-intervention data to conduct the evaluation. We follow a similar approach (not reported) and find no evidence of an enduring deterrent effect. If anything, the (equivocal) evidence points in the opposite direction.

[46] This can be difficult to judge, however, since several studies do not report standard errors or confidence intervals. Given that sample sizes are often under 100 or even under 30, it seems reasonable to assume that the confidence intervals include displacement effects.

prevalent in dense networks with a high chance of outcomes or even treatments spilling over to close units.

Flexibility in measuring spillovers is also crucial, and we illustrate how this can be a design-based choice, regardless of the inference method used. In Bogotá we find evidence of spillovers in a catchment area considerably wider than the usual catchment area, which if true could mean that the aggregate effect of displacement is considerably greater. Continuous rates of decay impose a fair degree of structure on the nature of the spillover, which is fine if that structure is well-understood. We illustrate a nonparametric alternative.

Lessons for crime prevention and state building From the perspective of crime and violence reduction, our results are consistent with a tenet of criminology: that crime and violence are highly concentrated in specific places. But if crime is easily displaced, then targeting, coordinating, and concentrating resources in high-crime places may not be the right approach after all. Rather, it might be wiser to target the specific people who commit crimes or particular behaviors. Displacement may be inherently less likely than in place-based approaches. This is the spirit of focussed deterrence, which identifies the small group of people who commit serious crimes and use threats and incentives to keep them from offending (Kennedy, 2011). This is also the spirit of cognitive behavioral therapy, which fosters skills and norms of non-violent behavior in high-risk young adults (Heller et al., 2017; Blattman et al., 2017).

From the broader perspective of state building, the effort to build the last mile of the state in Bogotá has parallels to a broader set of cases. The tendency for people to elude the state, or simply run away, is as old as state coercion. Targeted state interventions simply create the illusion of local control. It may be that state coercion and state presence have to be much more general, and much more widely spread, in order to be effective. The urban crime and violence literature has pushed theory and interventions to a more and more micro level, but to be effective, interventions might have to be more broad-based and stronger in order to keep crime from getting pushed to nearby places. The monopoly of violence is necessarily broad, and order is inconsistent with an ungoverned periphery. Small-scale trials may have led us to the opposite conclusion. Larger scale investigations, which are sorely needed in the U.S. and more globally, provide more precise tests.

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2016). Clustering as a Design Problem. *Working paper*.

Abt, T. and C. Winship (2016). What Works in Reducing Community Violence: A Meta Review and Field Study for the Northern Triangle. Democracy International, Inc, USAID, Washingon, DC.

Apel, R. (2013). Sanctions, perceptions, and crime: Implications for criminal deterrence. *Journal of quantitative criminology 29*(1), 67–101.

Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston, and N. Singh (2017). The Efficient Deployment of Police Resources: Theory and New Evidence from a Randomized Drunk Driving Crackdown in India. *Working paper*.

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy 76*, 169–217.

Berman, E. and A. M. Matanock (2015). The Empiricists' Insurgency. *Annual Review of Political Science*.

Blanes i Vidal, J. and G. Mastrobuoni (2017). Police Patrols and Crime. *Working paper*.

Blattman, C., J. Jamison, and M. Sheridan (2017). Reducing crime and violence: Experimental evidence on cognitive behavioral therapy in Liberia. *American Economic Review 107*(4).

Braga, A. and B. J. Bond (2008). Policing crime and disorder hot spots: A randomized controlled trial. *Criminology 46*, 577–608.

Braga, A., A. V. Papachristos, and D. M. Hurreau (2012). An ex post factor evaluation framework for place-based police interventions. *Campbell Systematic Reviews 8*, 1–31.

Braga, A., D. Weisburd, E. Waring, L. Green Mazerolle, W. Spelman, and F. Gajewski (1999). Problem-oriented policing in violent crime places: A randomized controlled experiment. *Criminology 37*, 541–580.

Cassidy, T., G. Inglis, C. Wiysonge, and R. Matzopoulos (2014). A systematic review of the effects of poverty deconcentration and urban upgrading on youth violence. *Health and Place 26*, 78–87.

Chalfin, A. and J. McCrary (2017). Criminal Deterrence: A Review of the Literature. *Journal of Economic Literature 55*(1), 5–48.

Chalfin, A. and J. McCrary (forthcoming 2017b). Are US Cities Underpoliced?: Theory and Evidence. *Review of Economics and Statistics*.

Collazos, D., E. Garcia, D. Mejia, D. Ortega, and S. Tobon (2017). Hotspots policing in a high crime environment: An experimental evaluation in Medellin. *In progress*.

Di Tella, R. and E. Schargrodsky (2004, March). Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack. *American Economic Review 94*(1), 115–133.

Farrington, D. P. and B. C. Welsh (2008). Effects of improved street lighting on crime: a systematic review. *Campbell Systematic Reviews* (13), 59.

Ferraz, C., J. Monteiro, and B. Ottoni (2016). State Presence and Urban Violence: Evidence from Rio de Janeiro's Favelas. *Working Paper*.

Gerber, A. S. and D. P. Green (2012). *Field experiments: Design, analysis, and interpretation*. New York: WW Norton.

Groff, E. R., J. H. Ratcliffe, C. P. Haberman, E. T. Sorg, N. M. Joyce, and R. B. Taylor (2015). Does what police do at hot spots matter? The philadelphia policing tactics experiment. *Criminology 53*(1), 23–53.

Heller, S. B., A. K. Shah, J. Guryan, J. Ludwig, S. Mullainathan, and H. A. Pollack (2017). Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago. *Quarterly Journal of Economics 132*(1), 1–54.

Horvitz, D. G. and D. J. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association 47*(260), 663–685.

Kennedy, D. M. (2011). *Don't shoot: one man, a street fellowship, and the end of violence in inner-city America*. Bloomsbury Publishing USA.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Mazerolle, L. G., J. F. Prince, and J. Roehl (2000). Civil remedies and drug control: a randomized field trial in Oakland, CA. *Evaluation Review 24*, 212–241.

Police Executive Research Forum, (2008). Violent crime in America: What we know about hot spots enforcement. Technical report, Police Executive Research Forum, Washingon, DC.

Ratcliffe, J. H., T. Tangiguchi, E. R. Groff, and J. D. Wood (2011). The Philadelphia Foot Patrol Experiment: A Randomized Controlled Trial of Police Patrol Effectiveness in Violent Crime Hotspots. *Criminology 49*, 795–831.

Scott, J. C. (2014). *The art of not being governed: An anarchist history of upland Southeast Asia*. Yale University Press.

Sherman, L., M. Buerger, and P. Gartin (1989). Beyond dial-a-cop: A randomized test of repeat call policing (recap). Washington, D.C.: Crime Control Institute.

Sherman, L. and D. P. Rogan (1995). Deterrent effects of police raids on crack houses: A randomized, controlled experiment. *Justice Quarterly 12*(4), 755–781.

Sherman, L. and D. Weisburd (1995). Does Patrol Prevent Crime? The Minneapolish Hot Spots Experiment. In *Crime Prevention in the Urban Community*. Boston: Kluwer Law and Taxation Publishers.

Sherman, L., S. Williams, A. Barak, L. R. Strang, N. Wain, M. Slothower, and A. Norton (2014). An Integrated Theory of Hot Spots Patrol Strategy: Implementing Prevention by Scaling Up and Feeding Back. *Journal of Contemporary Criminal Justice 30*(2), 95–122.

Taylor, B., C. Koper, and D. Woods (2011). A randomized controlled trial of different policing strategies at hot spots of violent crime. *Journal of Experimental Criminology 7*, 149–181.

Telep, C., R. Mitchell, and D. Weisburd (2014). How Much Time Should Police Spend at Crime Hot Spots? Answers from a Police Agency Directed Randomized Field Trial in Sacramento, California. *Justice Quarterly 31*(5), 905–933.

Weisburd, D., A. A. Braga, E. R. Groff, and A. Wooditch (2017). Can Hot Spots Policing Reduce Crime in Urban Areas? an Agent-Based Simulation. *Criminology 55*(1), 137–173.

Weisburd, D. and C. Gill (2014). Block Randomized Trials at Places: Rethinking the Limitations of Small N Experiments. *Journal of Quantitative Criminology 30*(1), 97–112.

Weisburd, D. and L. Green (1995). Measuring Immediate Spatial Displacement: Methodological Issues and Problems. In *Crime and Place: Crime Prevention Studies*, pp. 349–359. Monsey, NY: Willow Tree Press.

Weisburd, D., D. Groff, and S. Yang (2012). *The Criminology of Place: Street Segments and Our Understanding of the Crime Problem*. New York: Oxford University Press.

Weisburd, D. and C. Telep (2016). Hot Spots Policing: What We Know and What We Need to Know. *Journal of Experimental Criminology 30*(2), 200–220.

Weisburd, D. L., L. a. Wyckoff, J. E. Eck, and J. Hinkle (2005). Does Crime Just Move Around the Corner? A Study of Displacement and Diffusion in Jersey City, NJ. *Criminology 44*(August), 549–592.

Wilson, J. and G. Kelling (1982). Broken windows: The police and neighborhood safety. *Atlantic Monthly March*, 29–38.

Appendix for online publication

A Analysis of the existing literature

A.1 Power analysis

The aggregate effects on crime are difficult to pinpoint because of the small size of most studies. Figure A.1 plots the systematically-reviewed studies by sample size and effect sizes, for both direct and spillover effects.[1] We calculate statistical power curves, representing the minimum effect size that we would expect to be able to detect with 80% confidence.[2] Note that even the largest studies do not exceed 50 or 100 treated hot spots, with a similarly modest number of spillover segments. The average effect size for direct hot spots treatment across the studies is 0.17 standard deviations, and 0.24 if statistically significant.[3] While covariate adjustment and blocking strategies could improve statistical power slightly, these would produce at best marginal gains in precision.

In Bogotá, the city tested two place-based security interventions on a scale large enough to identify direct treatment effects of 0.15 standard deviations, and spillovers as small as 0.02 standard deviations. We plot these in Figure A.1. For fairness in the comparison, we plot the power of our study measured also on the basis of sample size and the number of treated units.

A.2 Overview of existing studies

Table A.1 summarizes the studies included in Braga et al. (2012). We also include more recent studies to complement the analysis. Power curves in Figure A.1 include all randomized controlled trials in the table.

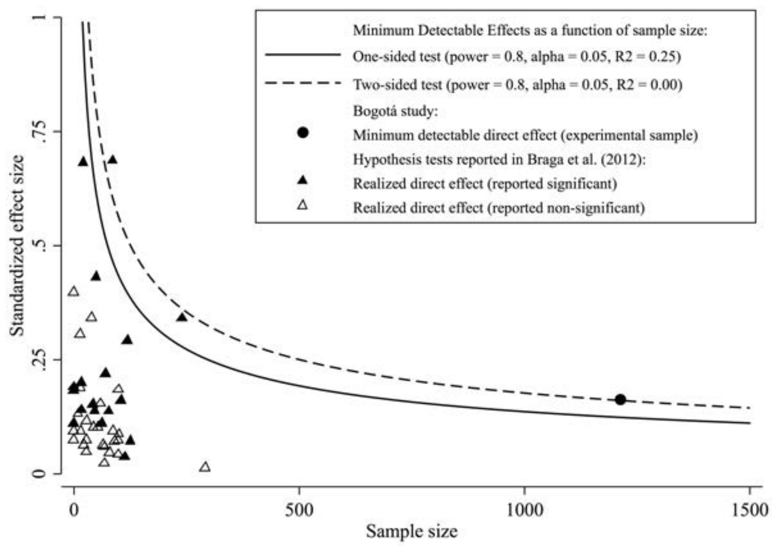
[1]The equations for the power curves are expected to be lower bounds of the actual power, as it could be increased using different randomization techniques as blocking by some specific characteristic of the units of analysis. Hence, some studies might have more power, given their sample size, than the corresponding value using the simple power formula. To make our study comparable to others, we also estimate our power using the formula rather than relying on our randomization approach. Another source of incomparability between studies could be the variation in outcomes within each experimental unit. As we show in Table A.1, some studies have units of analysis larger than a street segment as police beats. Some others have units of analysis smaller as specific addresses. In some cases, the main outcomes are calls for service, which might have more variation than crime reports in some contexts. Nonetheless, most of the studies focus on relatively small hot spots and we rely not only in crime reports but in an original survey of about 24,000 respondents. Hence, this source of incomparability should not be relevant.

[2]We generate the power curves assuming simple randomization and treatment assignment for half of the experimental sample. We acknowledge that some randomization procedures as blocking on pre-treatment characteristics could increase power (see for instance Gerber and Green, 2012; Weisburd and Gill, 2014), though the improvements may not be significant with small samples.

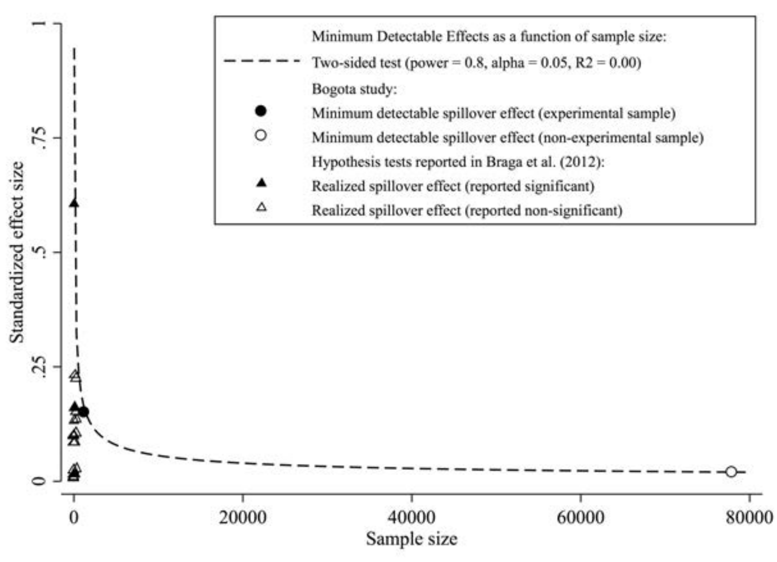
[3]We only report MDEs for studies for which it was possible to do so with the information in published papers. Generally study sizes are small, previous randomized controlled trials of intensive policing have sample sizes of 110 hot spots (55 treated) in Minneapolis (Sherman and Weisburd, 1995), 56 hot spots (28 treated) in Jersey City Weisburd and Green (1995), 24 hot spots (12 treated) in a different intervention in Jersey City (Braga et al., 1999), 207 hot spots (104 treated) in Kansas City (Sherman and Rogan, 1995), 100 hot spots (50 treated) in Oakland (Mazerolle et al., 2000), 34 hot spots (17 treated) in Lowell (Braga and Bond, 2008), 83 hot spots (21 treated with police patrols and 22 with problem oriented policing) in Jacksonville (Taylor et al., 2011), 120 hot spots (60 treated) in Philadelphia (Ratcliffe et al., 2011), and 42 hot spots (21 treated) in Sacramento (Telep et al., 2014). Interestingly, the first hot spots study was conducted in Minneapolis in 1989 and had a larger sample size with 250 residential addresses of which 125 were assigned to treatment and 250 commercial addressees of which also 125 were assigned to treatment Sherman et al. (1989). One of the only other large studies, by a subset of this paper's author's, is in the Colombian city of Medellín, with 384 of 967 hot spots treated Collazos et al. (2017). Even non-experimental sample sizes have been fairly small. Di Tella and Schargrodsky (2004), for instance, examined the effects of 37 police-protected religious institutions in Buenos Aires.

Figure A.1: Statistical power in the intensive policing literature

(a) Direct and spillover effects within the experimental sample of hot spots



(b) Spillover effects into "non-hot spots" proximate to the experimental sample



Notes: The figure depicts minimum detectable effects and realized effect sizes as a function of sample size. The vertical axis is in standard deviation units and measures minimum detectable effects for power curves and realize effect sizes for previous studies, and the horizontal axis measures sample size. The equations for power curves are $y = m \times 2\sqrt{\frac{1-R^2}{x}}$, where y is the standardized effect size, x is the sample size, and m is a multiple relating the standard deviation to the effect size. This multiple is 2.49 for one sided tests and 2.80 for two sided. Triangles represent a hypothesis test from previous studies and circles represent the minimum detectable effects in our study.

Table A.1: Review of previous literature on hot spots policing

Study and reference	Main characteristics	Technical details	Spillover analysis
Minneapolis Hot spots (Minneapolis, MN). Sherman, L., & Weisburd, D. (1995). General deterrent effects of police patrol in crime hot spots: A randomized controlled trial. Justice Quarterly 12, 625-648.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 12 months. *Main outcome:* Calls for service.	*Hot spot definition:* Address clusters identified using the number of calls for service. *Experimental units:* 110; 55 treated with intensive patrolling, 55 controls. *Randomization procedure:* Hot spots were assigned to five blocks based on hard crime call frequencies. Then randomized treatment within each block.	No analysis on spillovers.
Jersey City DMAP (Jersey City, NJ). Weisburd, D., & Green, L. (1995). Policing drug hot Spots: The Jersey City DMAP experiment. Justice Quarterly 12, 711-36.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 15 months. *Main outcome:* Calls for service.	*Hot spot definition:* Intersection areas identified using number of drug-related calls for service and narcotics arrests. *Experimental units:* 56; 28 treated with intensive patrolling (focussed on drugs), 28 controls. *Randomization procedure:* Hot spots were assigned to four blocks based on call frequencies and arrests. Then randomized treatment within each block.	*Method:* Two block catchment areas surrounding treatment and control hot spots.
Kansas City Crack House Raids (Kansas City, KS). Sherman, L., & Rogan, D. (1995). Deterrent effects of police raids on crack houses: A randomized controlled experiment. Justice Quarterly 12, 755-82.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 1 day (1 raid per hot spot). *Main outcome:* Calls for service.	*Hot spot definition:* Blocks identified using calls for service and court authorized raids. *Experimental units:* 207; 104 treated with police raids, 103 controls. *Randomization procedure:* Random assignment of treatment using the whole sample.	No analysis on spillovers.
Jersey City POP at violent places (Jersey City, NJ). Braga, A., Weisburd, D., Waring, E., Mazerolle, L.G., Spelman, W., & Gajewski, F. (1999). Problem-oriented policing in violent crime places: A randomized controlled experiment. Criminology 37, 541-80.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 16 months. *Main outcome:* Calls for service, crime reports, arrests.	*Hot spot definition:* Blocks identified using calls for service and court authorized raids. *Experimental units:* 24; 12 treated with problem oriented policing, 12 controls. *Randomization procedure:* Hot spots were matched in couples based on qualitative and quantitative assessments. Then randomized treatment within couples.	*Method:* Two block catchment areas surrounding treatment and control hot spots. Selected hot spots were cleared so final units were separate.

Notes: Continued on following page.

Review of previous literature on hot spots policing (continued)

Study and reference	Main characteristics	Technical details	Spillover analysis
Oakland Beat Health Program (Oakland, CA). Mazerolle, L., Price, J., & Roehl, J. (2000). Civil remedies and drug control: a randomized field trial in Oakland, California. Evaluation Review, 24, 212 – 241.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 5.5 months. *Main outcome:* Calls for service.	*Hot spot definition:* Street blocks referred as having drug or blight problems. *Experimental units:* 100; 50 treated with drug-related civil remedies, 50 controls. *Randomization procedure:* Random allocation blocking by economic use of land: residential and commercial	*Method:* 500 feet (about 150m) catchment areas surrounding treatment and control hot spots
Lowell Policing Crime and Disorder Hot Spots (Lowell, MA). Braga, A., & Bond, B. (2008). Policing crime and disorder hot spots: A randomized controlled trial. Criminology, 46 (3): 577 – 608.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 12 months. *Main outcome:* Calls for service.	*Hot spot definition:* Polygons built using spatial analysis of crime and disorder calls for service. *Experimental units:* 34; 17 treated with problem oriented policing, 17 controls. *Randomization procedure:* Hotspots were matched in couples based on qualitative and quantitative assessments. Then randomized treatment per couple.	*Method:* Two block catchment areas surrounding treatment and control hot spots. All hot spots were cleared so they included a two block catchment area to analyze spillovers.
Jacksonville Policing Violent Crime Hot Spots (Jacksonville, FL). Taylor, B., Koper, C., & Woods, D. (2011). A randomized controlled trial of different policing strategies at hot spots of violent crime. Journal of Experimental Criminology 7, 149-181.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 3 months. *Main outcome:* Calls for service and crime reports.	*Hot spot definition:* Land parcels built using spatial analysis of crime. Avg. hot spot size 0.02 sq. miles. The researchers revised locations so that each hot spot was at least one block away from any other. *Experimental units:* 83; 22 treated with problem oriented policing, 21 treated with intensive patrolling, 40 controls. *Randomization procedure:* Hot spots were arranged in four blocks according to violent crime reports. Then randomized each of the three conditions within blocks.	Method: 500 feet (about 150m) catchment areas surrounding treatment and control hot spots. All hot spots were cleared so that no hot spot was within a range of one block from another.

Notes: Continued on following page.

Review of previous literature on hot spots policing (continued)

Study and reference	Main characteristics	Technical details	Spillover analysis
Philadelphia Foot Patrol Program (Philadelphia, PA). Ratcliffe, J., Taniguchi, T., Groff, E., & Wood, J. (2011). The Philadelphia foot patrol experiment: A randomized controlled trial of police patrol effectiveness in violentcrime hot spots. Criminology 49 (3), 795-831.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 4 months. *Main outcome:* Crime reports.	*Hot spot definition:* Patrol beats identified using spatial analysis of violent crimes, validated with the Police Department. *Experimental units:* 120; 60 treated with intensive patrolling, 60 controls. *Randomization procedure:* Hot spots were ranked based on violent crime reports and matched in couples. Then randomized treatment within couples.	*Method:* Weighted displacement quotient with 2 block catchment areas.
Minneapolis RECAP (Minneapolis, MN). Sherman, L., Buerger, M., & Gartin, P. (1989). Beyond dial-a-cop: A randomized test of Repeat Call Policing (RECAP). Washington, DC: Crime Control Institute.	Randomized controlled trial included in Braga et al. (2012). *Intervention period:* 12 months. *Main outcome:* Calls for service.	*Hot spot definition:* Addresses identified using the frequency of calls for service. *Experimental units:* 500. 250 commercial units 125 treated with problem oriented policing and 125 controls; and 250 residential units with 125 treated with problem oriented policing and 125 controls. *Randomization procedure:* Random allocation within each group of experimental units	No analysis on spillovers.
Philadelphia Policing Tactics (Philadelphia, PA). Groff, E. R., Ratcliffe, J. H., Haberman, C. P., Sorg, E. T., Joyce, N. M., & Taylor, R. B. (2015). Does what police do at hot spots matter? The Philadelphia policing tactics experiment. Criminology, 53(1), 23–53.	Randomized controlled trial. *Intervention period:* About 7 months for problem oriented policing, 3 months for foot patrols and 7 months for the offender focused intervention. *Main outcome:* Crime reports.	*Hot spot definition:* Patrol beats identified using spatial analysis of violent crimes and validated with the Police Department. Average size was 0.044 sq. miles. *Experimental units:* 20 treated with problem oriented policing, 7 controls; 27 with 20 treated with foot patrols, 20 controls; and 27 with 20 treated with offender focused interventions and 7 controls. *Randomization procedure:* Blocked by police technique suitability according to a qualitative assessment by the Police Department.	*Method:* Weighted displacement quotient with 2 block catchment areas.

Notes: Continued on following page.

Review of previous literature on hot spots policing (continued)

Study and reference	Main characteristics	Technical details	Spillover analysis
Port St. Lucie Offender Focused Intervention (Port St. Lucie, FL). Santos, R. B., & Santos, R. G. (2016). Offender-focused police intervention in residential burglary and theft from vehicle hot spots: a partially blocked randomized control trial. Journal of Experimental Criminology, 1–30.	Randomized controlled trial. *Intervention period:* 9 months. *Main outcome:* Crime reports and arrests.	*Hot spot definition:* Aggregated census blocks identified using a qualitative assessment of neighborhoods and reported crimes. Average size was 0.60 sq. miles. *Experimental units:* 48; 24 treated with an offender focused intervention, 24 controls. *Randomization procedure:* 3 blocks of irregular sizes grouped according to a ranking on the rate of crimes per identified offender. Half of each blocked was randomly assigned to treatment.	No analysis on spillovers.
New York Tactical Narcotics Team (New York, NY). Sviridoff, M., Sadd, S., Curtis, R., & Grinc, R. (1992). The neighborhood effects of street-level drug enforcement: tactical narcotics teams in New York. New York: Vera Institute of Justice.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 3 months. *Main outcome:* Crime reports.	*Hot spot definition:* Streets, intersections and sets of buildings. *Non-experimental units:* 2 clusters (precincts) were targeted with tactical narcotics teams (hot spots within each precinct). *Approach:* Targeted hot spots matched with similar hot spots in a different precinct.	No analysis on spillovers.
St. Louis POP in 3 Drug Areas (St. Louis, MO). Hope, T. (1994). Problem-oriented policing and drug market locations: Three case studies. Crime Prevention Studies 2, 5-32.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 9 months. *Main outcome:* Calls for service.	*Hot spot definition:* Addresses with drug sales identified. *Non-experimental units:* 3 clusters targeted with problem oriented policing. *Approach:* Hot spot addresses were compared to other addresses on the same blocks and other blocks in surrounding areas.	*Method:* Calls for service in targeted addresses compared to calls for service in addresses at the same and surrounding blocks.
Kansas City Gun Project (Kansas City, KS). Sherman, L., & Rogan, D. (1995a). Effects of gun seizures on gun violence: 'Hot spots' patrol in Kansas City. Justice Quarterly 12, 673-694.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 7 months. *Main outcome:* Crime reports.	*Hot spot definition:* Police beats of 8 by 10 blocks. *Non-experimental units:* 1 cluster targeted with intensive enforcement on possession of firearms. *Approach:* The targeted beat was matched to a control beat according to the level of reported shootings.	*Method:* Time series analysis in 7 contiguous beats.

Notes: Continued on following page.

Review of previous literature on hot spots policing (continued)

Study and reference	Main characteristics	Technical details	Spillover analysis
Beenleigh Calls for Service (Beenleigh, AUS). Criminal Justice Commission. (1998). Beenleigh calls for service project: Evaluation report. Brisbane, Queensland, AUS: Criminal Justice Commission.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 6 months. *Main outcome:* Calls for service.	*Hot spot definition:* Suburb with addresses with large number of calls for service. *Non-experimental units:* 1 cluster targeted with problem oriented policing. *Approach:* Trends in calls for service in the targeted suburb.	No analysis on spillovers.
Houston Targeted Beat Program (Houston, TX). Caeti, T. (1999). Houston's targeted beat program: A quasi-experimental test of police patrol strategies. Ph.D. diss., Sam Houston State University. Ann Arbor, MI: University Microfilms International.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 24 months. *Main outcome:* Crime reports.	*Hot spot definition:* Beats with highest reported crime. *Non-experimental units:* 3 hot spots targeted with highly visible patrols, 3 targeted with zero tolerance patrols, 1 targeted with problem oriented policing. *Approach:* Targeted beats were matched to non-contiguous beats.	*Method:* Time series analysis in contiguous beats.
Pittsburgh Police Raids (Pittsburg, PA). Cohen, J., Gorr, W., & Singh, P. (2003). Estimating intervention effects in varying risk settings: Do police raids reduce illegal drug dealing at nuisance bars? Criminology, 41 (2): 257 – 292.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 5 months. *Main outcome:* Calls for service.	*Hot spot definition:* Nuisance bar areas with 200m radius. *Non-experimental units:* 37 areas targeted with police raids. *Approach:* Targeted bar areas were compared to non-nuisance 40 bar areas.	No analysis on spillovers.
Buenos Aires Police after Terrorist Attack. DiTella, R., & Schargrodsky, E. 2004. Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. American Economic Review 94, 115 – 133.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 5 months. *Main outcome:* Crime reports.	*Hot spot definition:* Street blocks with Jewish centers that received increased police presence after a terrorist attack. *Non-experimental units:* 1 cluster targeted with intensive enforcement on possession of firearms. *Approach:* Targeted street blocks compared with >800 other blocks.	*Method:* One and two blocks catchment areas surrounding targeted areas.

Notes: Continued on following page.

Review of previous literature on hot spots policing (continued)

Study and reference	Main characteristics	Technical details	Spillover analysis
Philadelphia Drug Corners Crackdowns (Philadelphia, PA). Lawton, B., Taylor, R., & Luongo, A. (2005). Police officers on drug corners in Philadelphia, drug crime, and violent crime: Intended, diffusion, and displacement impacts. Justice Quarterly 22, 427 – 451.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 4.5 months. *Main outcome:* Crime reports.	*Hot spot definition:* High activity drug locations with an area of 0.1 miles. *Non-experimental units:* 214 locations targeted with police crackdowns. *Approach:* Targeted locations matched with a sample of 73 other locations.	*Method:* Adjoining areas (between 0 and 0.1 miles from target sites) were compared with comparison areas (more than 0.2 miles away from target sites).
Jersey City Displacement and Diffusion Study (Jersey City, NJ). Weisburd, D., Wyckoff, L., Ready, J., Eck, J., Hinkle, J., and Gajewski, F. (2006). Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits. Criminology 44, 549 – 592.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 6 months. *Main outcome:* Crime reports.	*Hot spot definition:* Two areas comprising 81 and 88 street segments were identified according to drug sales and prostitution, respectively. *Non-experimental units:* some street segments in each area were targeted with problem oriented policing, other streets were ex-post assigned to a short range displacement area and a long range displacement area. *Approach:* Trends in prostitution and drug events were observed in targeted and displacement areas.	*Method:* Trends in catchment areas.
Boston Safe Street Program (Boston, MA). Braga, A. A., Hureau, D. M., & Papachristos, A. V. (2012). An Ex Post Facto Evaluation Framework for Place-Based Police Interventions, Evaluation Review 35(6), 592–626.	Non-experimental study included in Braga et al. (2012). *Intervention period:* 36 months. *Main outcome:* Crime reports.	*Hot spot definition:* Streets with large number of reported violent crimes. *Non-experimental units:* 13 clusters targeted with Safe Street Program. *Approach:* Street segments within the boundaries of the targeted areas were matched to street segments outside the areas.	*Method:* Two block catchment areas surrounding targeted areas were compared to catchment areas of matched areas.

Notes: This table summarizes the studies included in Braga et al. (2012) with additional, more recent studies (noted in the main characteristics column.

B Additional data and design details

B.1 Construction of the experimental sample

To identify the initial pool of streets (the 2% highest crime segments) we drew on geolocated National Police data. We constructed a geo-fence of 40m around each segment and assigned a reported crime to that segment whenever it fell within its geo-fence. We ranked segments based on a weighted sum of the crimes of most concern to the Mayor's office: homicides, assaults, robberies, car theft and motorcycle theft.[4] If there was a crime within two or more geo-fences, we assigned the crime to the closest segment using linear distances. Thus if a crime occurred in a public park, it would be assigned to the nearest segment. There are some missing data, especially in the first two years of the data, when about a quarter of reported crimes could not be geo-coded because of deficiencies in the address data. From 2014 onwards, the crime data come with a geographically coordinate, but in some cases these coordinates do not fall within any 40 meter fence and were therefore not assigned to a segment. It was also possible for crime locations to be mis-recorded by the police or citizen

Notably, most of the streets added by police had no reported crimes in the 2012–15 police database. The police nonetheless perceived them to be hot spots because they were known as areas of unreported crime such as pickpocketing, drug sales, or muggings. In eliminating streets, the police said that they dropped segments that they suspected had erroneous crime levels because of their location. For instance, streets close to a police or CAI station, a bus station, or a hospital might have too many crimes in the administrative data, because they were incorrectly designated as the crime site.

B.2 Patrolling time

Figure B.1 presents the evolution of average daily patrolling time for the pre-treatment and treatment periods, as well as different groups of streets: treatment, controls (all) and non-experimental.

Our estimates of average daily patrolling time are lower in the pre-treatment period because of data quality. During the pre-treatment period not all police patrols had GPS devices and some were working irregularly as the equipment was being piloted. During the treatment period there were also windows of intermittence. These malfunctioning periods, however, affected all streets equally.[5] Even though we cannot compare average daily patrolling time between the pre-treatment and treatment periods directly, the figures show that average patrolling time in control streets is between two and three times as much as that for non-experimental streets. This is true for both periods and especially for time windows where the GPS devices seemed to be working better.[6]

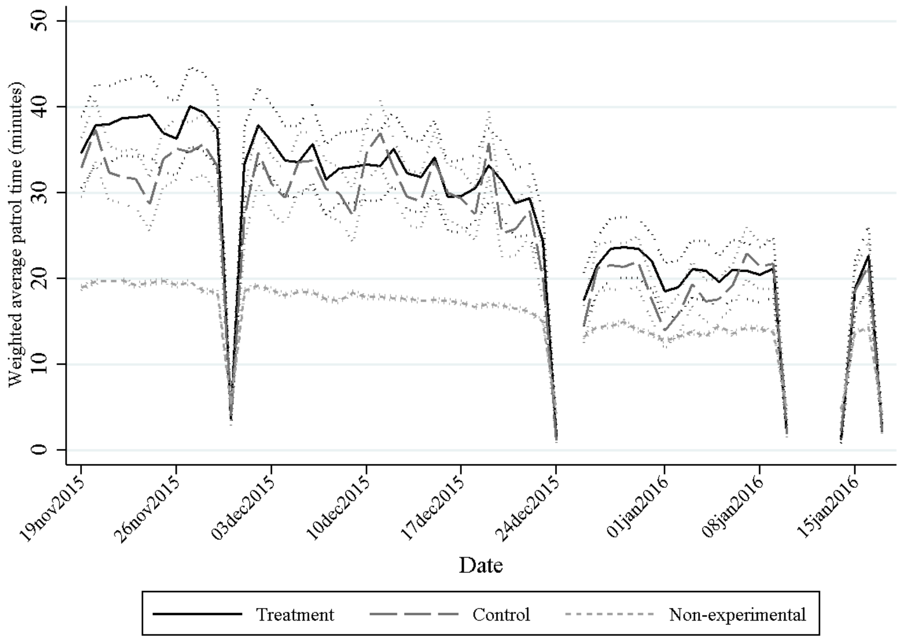
[4]We based crime weights on the average prison sentence according to Colombian law, which proxy for the social costs of crime. For the aggregate crime index, weights are: 0.300 for homicides, 0.112 for assaults, 0.116 for theft from person and 0.221 for car and motorcycle theft. The weight for homicides was cut by half in order to avoid every segment with one homicide in the past four years to become a hot spot. At the Mayor's office direction, we did not use data on family violence, sexual assault, shoplifting, threats, and other lower frequency crimes to determine hot spots. A focus on homicides, vehicle theft, and robbery is also consistent with evidence from U.S. cities that these crimes respond most elastically to increased police presence (Chalfin and McCrary, 017b).

[5]We estimated patrolling time using the time stamp of the GPS pings sent by every device. In the easiest cases, several sequential pings were received from the area of 40m surrounding a segment. In this case, we took the first ping as the entry time and the last as the exit time, and computed the patrolling time for an entry. Then, we aggregated entries to measure daily patrolling times. However, because of malfunctioning units, there were several cases in which irregular and largely separated pings were sent by a device. To account for these situations, we top-coded each entry up to the duration of the shift (starting with the entry time). We also drop days with missing data, as it was more likely that the device was not working than the street was not patrolled at all during the day. We discussed these adjustments with the police to ensure we were making a correct approximation of daily patrolling times. The police reported that most cases were due to software updates in all devices. For instance, to update the operating system or the software for background checks.

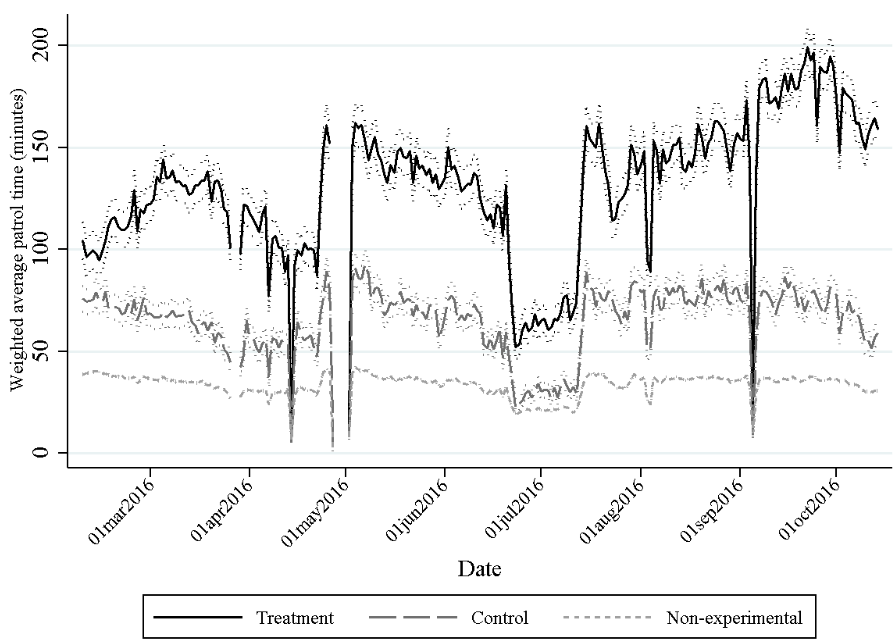
[6]For our estimates, we follow each GPS device chronologically, thus we track the moment at which the device enters a street and when does it leave. We made two assumptions to estimate patrolling time: (i) If we see only one GPS ping in a street and then the device moves to other streets, we impute 1 minute of patrolling time (assuming the patrol just traversed the street). (ii) If we see a device entering a street and the next ping from the same device is many hours ahead in the same street, we count until the end of the shift (assuming the device was maybe left there, but in any case the maximum patrolling time should go as much as the end of the shift).

Figure B.1: Evolution of patrolling time in the pre-treatment and treatment periods

(a) Pre-treatment period (November 2015 – January 2016)

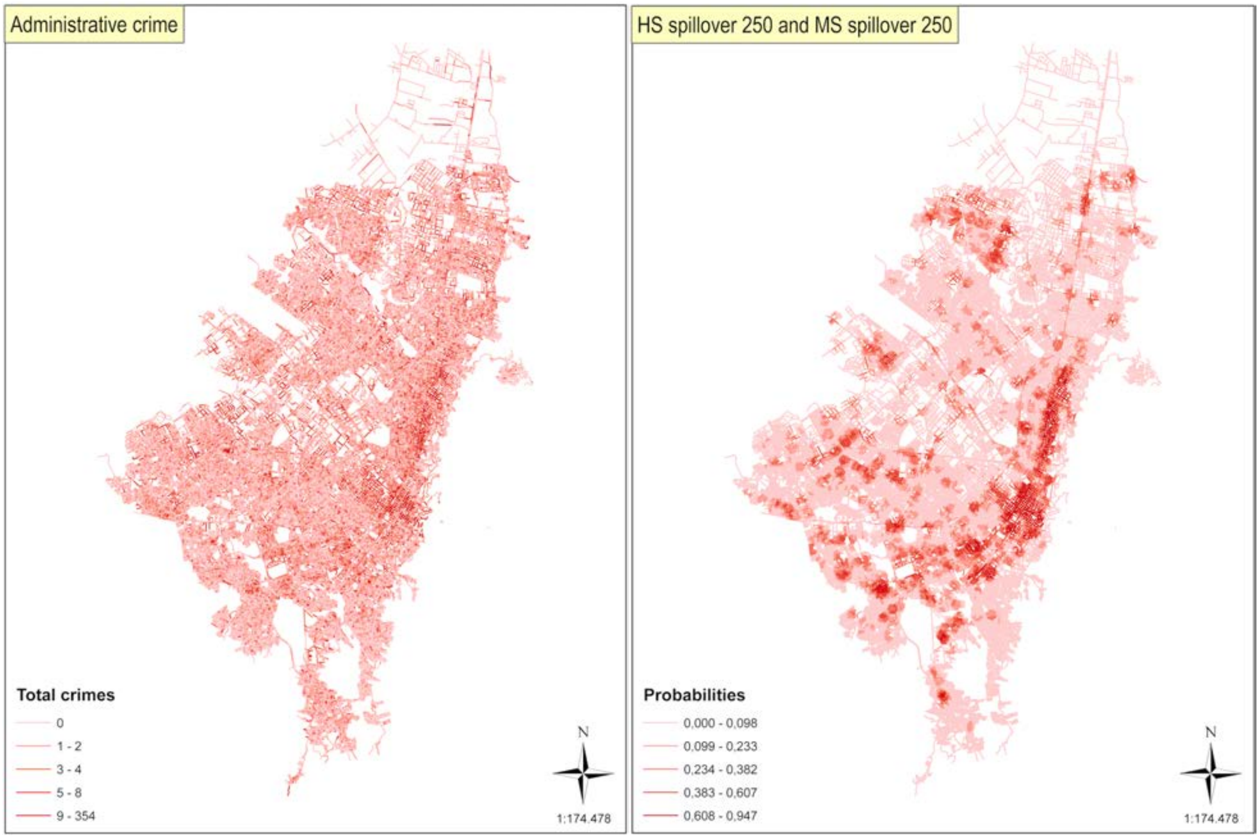


(b) Treatment period (February 2016 – October 2016)



Notes: The figures present estimates of the average daily patrolling time for the pre-treatment period: November 19, 2015 through January 14, 2016, and the treatment period: February 9, 2016 through October 14, 2016.

Figure C.1: Maps of baseline crime and probability of being spillover <250m to both interventions



Notes: This figure displays two maps of Bogota. In the first map, we display baseline administrative crime from 2012 to 2015 at the street-segment level. In the second map, we display each segment's probability of being within 250m of segments assigned to receive both interventions.

C Inverse probability weighting

Our randomization procedure gives segments variable probabilities of being in each of the treatment conditions. This is especially true for segments in our non-experimental sample. For example, non-experimental segments in relatively safer areas of Bogota have a zero percent chance of being a spillover for either treatment since there are no experimental units in those neighborhoods.

Figure C.1 compares two maps. The first map displays the number of baseline administrative crimes between 2012 and 2015 for each segment, while the second one displays each segment's probability of being within 250m of hotspots receiving hotspot policing and municipal services (based on 1,000 randomizations). In areas with lots of crime, non-experimental units have a higher probability of being a <250m spillover because they are located in areas with more hotspots (experimental units). In areas like the south of Bogota, however, many segments have no a zero probability of being a <250 spillover because there are no hotspots present. Thus a simple spillover vs. control comparison will lead to biased estimates on the effect of crime because the outcome (crime) is correlated with treatment assignment. In order to deal with this issue, we must use inverse probability weights and (in the case of the non-experimental units) omit units with a zero probability of being a spillover (so they are always controls) or being a control (so they are always spillovers).

In table C.1 we display the average bias associated with the use of inverse probability weights for our design. The top half shows the bias for the experimental sample while the bottom half shows the bias for the non-experimental sample. There are 1,916 units in the experimental sample, so the asymptotic requirement is unlikely to be met, leading to large biases associated with the design. By contrast, we have many more non-experimental units, which gives us much smaller biases.

Table C.1: IPW bias

		Experimental sample							
		Treatment effect				Spillover effect			
Outcome	Interaction included? (1)	Intensive policing (2)	Municipal services (3)	Interaction effect (4)	Both (2+3+4) (5)	Intensive policing (6)	Municipal services (7)	Interaction effect (8)	Both (6+7+8) (9)
Insecurity index, z-score (+ more insecure)	No	0.040	0.051		0.091	0.067	0.084		0.151
	Yes	0.027	0.038	0.031	0.095	0.083	0.096	-0.039	0.140
Perceived risk index, z-score (+ riskier)	No	0.037	0.039		0.076	0.064	0.072		0.135
	Yes	0.024	0.026	0.031	0.081	0.078	0.083	-0.036	0.125
Crime index, z-score (+ more crime)	No	0.030	0.045		0.075	0.049	0.068		0.117
	Yes	0.021	0.037	0.020	0.078	0.061	0.078	-0.030	0.108
Perceived & actual incidence of crime, z-score (survey)	No	0.035	0.040		0.075	0.062	0.061		0.123
	Yes	0.021	0.026	0.032	0.080	0.078	0.073	-0.040	0.111
# crimes reported to police on street segment (admin)	No	0.013	0.044		0.057	0.011	0.065		0.077
	Yes	0.017	0.048	-0.009	0.055	0.012	0.067	-0.003	0.076

		Non-experimental sample			
		Spillover effect			
	Interaction included? (1)	Intensive policing (6)	Municipal services (7)	Interaction effect (8)	Both (6+7+8) (9)
Insecurity index, z-score (+ more insecure)	No	0.003	0.010		0.013
	Yes	0.002	0.009	0.001	0.012
Perceived risk index, z-score (+ riskier)	No	0.002	-0.006		-0.004
	Yes	-0.003	-0.012	0.010	-0.005
Crime index, z-score (+ more crime)	No	0.003	0.023		0.026
	Yes	0.007	0.027	-0.008	0.026
Perceived & actual incidence of crime, z-score (survey)	No	-0.001	0.009		0.008
	Yes	0.004	0.014	-0.010	0.008
# crimes reported to police on street segment (admin)	No	0.009	-0.010		-0.001
	Yes	0.013	-0.007	-0.006	0.000

Notes: The table displays the average bias associated with the use of inverse probability weights. The first part presents the average bias for the experimental sample. The second part presents the average bias for the non-experimental sample.

D Departures from the pre-analysis plan

The estimation procedure used in this paper is slightly different from the ones we described in our pre-analysis plan (PAP).[7] In this section, we document the reasons why it was appropriate to switch estimation strategies.

Our pre-specified estimation strategy (see page 17 of the PAP) would use pairwise regressions to estimate the direct and spillover effects of the intervention. Let us assume we wanted to estimate the effects of the hot spot policing treatment given one level of spillovers, so our possible experimental conditions are: treated by hotspot policing T_H, <250m of a unit treated with hotspot policing S_H, and >250m away from a unit receiving hotspot policing (C_H, the control group). Our pre-analysis plan says we would run the following WLS regression:

$$Y_{sqp} = \beta_0 + \theta_H * T_H + \emptyset * X_{sqp} + \gamma_p + \varepsilon_{sqp} \tag{1}$$

Our weights are determined by the probability of being either in T_H, S_H, or C_H (for example, if a street is in S_H, its weight is $\frac{1}{\Pr(S_H)}$). Furthermore, we restrict the regressions to (i) segments only in T_H or C_H, and (ii) segments with a non-zero probabilities of being in T_H and C_H (i.e. $0 < \Pr(T_H) < 1 \bigcup 0 < \Pr(C_H) < 1$). The coefficient of interest is θ_H, which represents the ITT estimate of receiving the hot spot policing treatment on outcome Y relative to segments greater than 250m away from any treated hotspot.

This pairwise regression is incorrect because it fails to recognize the complexity of our design. We test both hot spot policing and municipal services in a factorial design, so probability weights need to be determined by the *joint* probability of hot spot policing and municipal service assignment, not just assignment to one of the treatments. Failure to account for the joint probability can mix up effects between each of the interventions. For example, if segments treated by hot spot policing have a higher chance than hot spot policing control segments to be inner spillovers for municipal services, then θ_H in equation 1 will conflate the direct effect of hot spot policing and the spillover effect of municipal services.

This is exactly what we see in our design. In Table D.1, we show the distribution of treatment assignments for each intervention. Panel A shows that while segments in each hot spot policing block all have a similar proportion (~11%) of their segments receiving municipal services, segments treated with policing are more likely than segments >250m from treated policing segments to be spillover units for municipal services. In the case that there are spillover effects from municipal services, it will not be possible to use the pairwise regression detailed above to estimate just the effect of hot spot policing.

There are two changes we can make to the regressions outlined in the pre-analysis plan so that our empirical strategy is compatible with the realities of our factorial design. First, we can base our probability weights on the joint probability of assignment. Second, we can insert dummies for municipal service assignment into equation 1. Making these changes gives us the following regression:

$$Y_{sqp} = \beta_0 + \theta_H * T_H + \theta_M * T_M + \theta_H * S^M + \emptyset * X_{sqp} + \gamma_p + \varepsilon_{sqp} \tag{2}$$

Including an additional indicator for being a hot spot policing spillover in this regression allows us to estimate all four effects (direct effect of hot spot policing, direct effect of municipal services, spillover effect of hotspot policing, spillover effect of municipal services) in one regression. This corresponds to the constrained version of equation (1) in the main paper where $\beta_3 = 0$. Thus the regressions used in this paper correctly estimate the effects of our factorial design by using the correct inverse probability weights and estimating all the effects in the same regression.

Nevertheless, we display the pairwise regressions pre-specified for clarity purposes. Table D.2 displays the hot spots policing effect while table displays the municipal services effects. Meanwhile, table D.4 displays the interaction effects. Most of the differences for the treatment effects are coming from the use of different weights. In Table D.4 (where we use the same weights as in the main analysis), the results are very similar—the only difference is that we drop observations that are within 250m of either treatment, giving us less power.

[7] https://www.socialscienceregistry.org/trials/1156.

Table D.1: Distribution of assignments, by treatment

Panel A: Distribution of municipal service assignments					
			Municipal services assignment		
		Total	Treated	<250m	>250m
		(1)	(2)	(3)	(4)
Intensive	Treated	756	0.10	0.26	0.64
policing	<250m	705	0.10	0.40	0.50
assignment	>250m	458	0.11	0.15	0.74
		1919			
Panel B: Distribution of policing assignments					
			Hotspot policing assignment		
		Total	Treated	<250m	>250m
		(1)	(2)	(3)	(4)
Municipal	Treated	201	0.37	0.37	0.26
services	<250m	546	0.36	0.51	0.13
assignment	>250m	1172	0.41	0.30	0.29
		1919			

Notes: This table displays the distribution of treatment assignments for each intervention. Panel A depicts the proportion of streets assigned to the different treatment status on municipal services, within each treatment block for hot spot policing. Panel B depicts the proportion of streets assigned to the different treatment status on hot spots policing, within each treatment block for municipal services.

Table D.2: Hot spots policing impacts on insecurity, pre-specified regressions

	ITT of assignment to:				
	Accounting for 250-500m spillovers	Accounting for spillovers <250m			No spillovers
Dependent variable	HSP outer spillover (1)	HSP treated (2)	HSP inner spillover (experimental) (3)	HSP inner spillover (non-experimental) (4)	HSP treated (5)
Insecurity index, z-score (+ more insecure)	0.078	-0.106	-0.055	0.122	-0.063
	0.366	*0.291*	*0.816*	*0.223*	*0.193*
Perceived risk index, z-score (+ riskier)	-0.014	-0.110	-0.080	0.120	-0.067
	0.834	*0.232*	*0.515*	*0.303*	*0.151*
Crime index, z-score (+ more crime)	0.144	-0.067	-0.011	0.083	-0.039
	0.169	*0.520*	*0.902*	*0.300*	*0.445*
Perceived & actual incidence of crime, z-score	0.175	-0.087	-0.030	0.110	-0.045
	0.137	*0.355*	*0.883*	*0.330*	*0.346*
# crimes reported to police on street segment	0.050	-0.011	0.029	0.013	-0.016
	0.651	*0.943*	*0.760*	***0.091***	*0.861*

Notes: This table reports intent to treat (ITT) estimates of the effects of hotspot policing using the pre-specified regressions. Randomization inference p-values are italicized.

Table D.3: Municipal services impacts on insecurity, pre-specified regressions

	ITT of assignment to:				
	Accounting for 250-500m spillovers	Accounting for spillovers <250			No spillovers
Dependent variable	MS outer spillover	MS treated	MS inner spillover (experimental)	MS inner spillover (non-experimental)	MS treated
	(1)	(2)	(3)	(4)	(5)
Insecurity index, z-score (+ more insecure)	-0.097	-0.138	0.082	-0.092	-0.152
	0.471	*0.092*	*0.173*	*0.419*	***0.005***
Perceived risk index, z-score (+ riskier)	-0.020	-0.137	0.031	-0.144	-0.128
	0.879	*0.101*	*0.508*	*0.189*	***0.008***
Crime index, z-score (+ more crime)	-0.142	-0.092	0.105	-0.009	-0.126
	0.161	*0.267*	*0.102*	*0.917*	***0.018***
Perceived & actual incidence of crime, z-score	-0.060	-0.129	0.056	0.032	-0.129
	0.656	*0.149*	*0.355*	*0.758*	***0.023***
# crimes reported to police on street segment	-0.264	0.006	0.175	-0.017	-0.088
	0.057	*0.971*	*0.141*	***0.019***	*0.379*

Notes: This table reports intent to treat (ITT) estimates of the effects of municipal services using the pre-specified regressions. Randomization inference p-values are italicized.

Table D.4: Interaction impacts on insecurity, pre-specified regressions

	ITT of assignment to:					
	Accounting for spillovers <250m			No spillovers		
Dependent variable	HSP effect	MS effect	Interaction effect	HSP effect	MS effect	Interaction effect
	(2)	(3)	(4)	(5)	(6)	(7)
Insecurity index, z-score (+ more insecure)	-0.177	-0.179	0.058	-0.077	-0.074	-0.185
	0.152	*0.140*	*0.573*	*0.220*	*0.273*	*0.120*
Perceived risk index, z-score (+ riskier)	-0.191	-0.183	0.058	-0.087	-0.053	-0.139
	0.120	*0.112*	*0.567*	*0.164*	*0.414*	*0.226*
Crime index, z-score (+ more crime)	-0.104	-0.115	0.038	-0.042	-0.070	-0.170
	0.368	*0.360*	*0.706*	*0.528*	*0.298*	*0.179*
Perceived & actual incidence of crime, z-score	-0.147	-0.255	0.229	-0.057	-0.138	0.040
	0.250	*0.047*	*0.137*	*0.405*	*0.052*	*0.686*
# crimes reported to police on street segment	-0.003	0.182	-0.329	-0.004	0.076	-0.526
	0.940	*0.372*	*0.295*	*0.957*	*0.579*	*0.024*

Notes: This table reports intent to treat (ITT) estimates of the effects of both interventions using the pre-specified regressions. Randomization inference p-values are italicized.

Table E.1: Testing for spillovers: F-tests of weighted mean differences between control regions

Outcome	p-value from F-test of joint significance			
	Experimental sample (N = 1,919)		Non-experimental sample (N=77,848)	
	250–500m vs >500m regions	<250m vs >250m regions	250–500m vs >500m regions	<250m vs >250m regions
	(1)	(2)	(3)	(4)
A. Intensive policing				
Perceived risk	0.235	0.717		
Crime incidence	0.542	0.716		
# crimes reported to police	0.626	0.165	0.277	0.224
B. Municipal services				
Perceived risk	0.667	0.648		
Crime incidence	0.434	0.093		
# crimes reported to police	0.434	0.029	0.576	0.552

Notes: There are 4×7 experimental conditions, with means reported in Table (2). This table tests for mean differences iteratively, first between the >500 meter and 250–500 meter conditions, then between the <250 meter and >250 meter conditions. It does so for each intervention. For instance, to test for spillovers in the 250-500m spillover region from from municipal services, we calculate the mean differences between the four cells in column 3 of Table (2) and the adjoining cells in column 4. This table reports the p-value from the F-test of those four mean differences.

E Spillovers

E.1 Tests of spillovers

Table E.1 reports the p-values from our preferred, general test of spillovers. It takes the means for the 4×5 experimental conditions in Table 2 in the paper and tests for differences between pairs of columns (for municipal services) and pairs of rows (for intensive policing). Using our pre-specified threshold of p<0.1, we observe statistically significant spillovers with 250m for municipal services, but not in the 250-500m region. For intensive policing, however, none of the p-values are below 0.1. We see some indication of <250m spillovers from municipal services in one of the two outcomes (crime incidence), but spillovers are not statistically significant in the large non-experimental sample.

This is one reason why we see more statistically significant spillovers in Table 6. We should also have addressed how we would treat economically large spillovers around or below $p = 0.1$. Because the spillovers in Table E.1 are weak, there is a reasonable argument for calculating treatment effects ignoring spillovers. We report these in the next Appendix section.

E.2 Alternate spillover regions, including no spillovers case

The right spillover region is not clear cut. We use the tests reported in Table E.1 to determine our main specification. But an argument could be made for wider and narrower spillovers.

First, there is weak evidence of spillovers beyond 250m. Table E.2 reports the same aggregate analysis of impacts on officially reported crime as in Table 6, but accountingfor spillovers <250m and spillovers 250-500m. We generally do not see evidence of statistically significant spillovers.

Tables E.3 reports the coefficients on spillover effects <250m and 250-500m on our main indexes using survey data. It is akin to Table 8 in the paper, but reports only the coefficients on the spillover terms that come from estimating equation (1) for <250m and 250-500m indicators. We see some evidence of spillovers 250–500m in the crime index.

Second, one could also make a case for no spillovers, at least with our permissive p<0.1 threshold. Table E.4 estimates equation (1) without spillovers. Conventional standard errors clustered at the quadrant level now produce reliable estimates. Qualitatively we draw the same conclusions. The main change is that the direct effects of treatment on reducing crime are more statistically significant. Looking at panel (a), without

Table E.2: Estimated aggregate impacts of the interventions, accounting for spillovers <250m and spillovers 250-500m in the experimental and non-experimental samples

	Dependent variable: # of crimes reported to police on segment (administrative data)							
	No interaction between treatments				Interaction between treatments			
	Coeff.	RI p-value	# segments	Estimated total impact = (1) × (3)	Coeff.	RI p-value	# segments	Estimated total impact = (5) × (7)
Impacts of treatment	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Direct treatment effect								
Intensive policing	-0.040	0.814	756	-30.4	0.046	0.740	756	34.8
Municipal services	-0.144	0.434	201	-29.0	0.032	0.635	201	6.5
Both					-0.491	**0.018**	75	-36.9
Subtotal				-59.4				4.5
B. <250m spillover, experimental sample								
Intensive policing	0.129	0.365	705	91.1	0.201	0.207	705	142.0
Municipal services	0.088	0.289	546	47.9	0.184	0.105	546	100.2
Both					-0.337	**0.081**	281	-94.7
Subtotal				139.0				147.5
C. 250-500m spillover, experimental sample								
Intensive policing	-0.051	0.768	294	-15.0	-0.028	0.893	294	-8.2
Municipal services	-0.151	0.249	495	-74.6	-0.137	0.324	495	-67.6
Both					-0.070	0.723	102	-7.1
Subtotal				-89.6				-82.9
D. <250m spillover, non-experimental sample								
Intensive policing	0.008	0.363	45209	375.7	0.010	0.336	45209	457.0
Municipal services	-0.003	0.834	18538	-46.6	0.000	0.922	18538	1.6
Both					0.006	0.841	14323	78.8
Subtotal				329.0				537.4
E. 250-500m spillover, non-experimental sample								
Intensive policing	0.011	0.356	16828	178.4	0.008	0.421	16828	129.6
Municipal services	0.014	0.494	21149	298.6	0.011	0.613	21149	230.9
Both					0.006	0.762	5965	33.1
Subtotal				476.9				393.6
Net increase (decrease) in crime				796.0				1,000.1
			95% CI	(-1548, 3565)			95% CI	(-1710, 3590)
			90% CI	(-1157, 3149)			90% CI	(-1224, 3198)

Notes: This table presents the aggregate effect calculation for both interventions assuming spillovers <250m and spillover 250-500m. Columns 1–4 refer to the non-interacted results (equation (1) under the constraint that $\beta_3 = 0$ and $\lambda_3 = 0$) while columns 5–8 refer to the interacted results (equation (1) with no constraints). Columns 1 and 5 display the bias-adjusted treatment effect while columns 2 and 6 display RI p-values. Columns 3 and 7 display the number of units in each group. Columns 4 and 8 display the product of the bias-adjusted treatment effect and the number of units in each group. The confidence interval on the bottom of the table is constructed using randomization inference. First we create a fake schedule of potential outcomes for each observation by adding or subtracting RI-adjusted treatment or spillover effects. This process gives us a potential outcome for each unit depending on its treatment assignment. Second, we simulate a randomization and take the potential outcome associated with the treatment assignment of the new randomization. Third we estimate treatment and spillover effects using this new outcome and apply the RI bias adjustment from our main set of results. Fourth, we multiply these bias-adjusted treatment effects by the number of segments in each group, and sum across both the experimental and non-experimental samples to get the aggregate effect. We repeat steps two through four 1,000 times to get the distribution of the test statistic, which is roughly centered on the actual number of deterred crimes. The 2.5 and 97.5 percentiles of this distribution give us the 95% confidence interval. $p < .1$ in bold.

Table E.3: Spillover effects in the experimental sample, accounting for spillovers within 250m, with p-values from randomization inference (N=1,916)

		Impact of spillovers <250m:				Impact of spillovers 250-500m:			
Dependent variable	Control mean	Any intensive policing	Any municipal services	Both interven-tions	Sum of (2), (3), and (4)	Any intensive policing	Any municipal services	Both interven-tions	Sum of (6), (7), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Insecurity index, z-score (+ more insecure)	-0.242	-0.057	-0.183	-0.107	-0.347	0.097	0.088	-0.228	-0.043
		0.858	*0.596*	*0.552*	*0.752*	*0.184*	*0.168*	***0.024***	*0.624*
Perceived risk index, z-score (+ riskier)	-0.098	-0.106	-0.171	-0.058	-0.335	0.001	0.027	-0.138	-0.110
		0.889	*0.698*	*0.811*	*0.783*	*0.515*	*0.370*	*0.120*	*0.786*
Crime index, z-score (+ more crime)	-0.305	0.010	-0.133	-0.120	-0.243	0.162	0.119	-0.242	0.039
		0.545	*0.552*	*0.443*	*0.666*	***0.074***	***0.078***	***0.019***	*0.377*
Perceived & actual incidence of crime, z-score	-0.149	-0.010	-0.203	0.091	-0.122	0.120	0.069	-0.160	0.029
		0.622	*0.277*	*0.296*	*0.841*	*0.160*	*0.226*	*0.110*	*0.432*
# crimes reported to police on street segment	0.473	0.046	0.032	-0.491	-0.413	0.201	0.184	-0.337	0.048
		0.740	*0.635*	*0.018*	***0.077***	*0.207*	*0.105*	***0.081***	*0.539*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports spillover estimates of equation (1) including two different spillover regions: 0-250m and 250-500m. Spillovers <250m are reported in columns 2-5 and spillovers 250-500m are reported in columns 6-9. We estimate the coefficients via a WLS regression of each outcome on spillover indicators, police station (block) fixed effects, and baseline covariates. The measures of perceived risk, perceived incidence of crime, and proportion reporting crime come from our citizen survey, and the # of crimes reported to the police come from police administrative data.

the interaction, intensive policing reduces the overall index of insecurity by about 0.12 standard deviations, and municipal services reduces it by about 0.16 standard deviations. Both perceived risk and crime incidence fall but, for intensive policing at least, the fall in crime incidence is not statistically significant. As before, we see the largest and most statistically robust impacts of state presence in the segments that received both interventions. Looking at the overall insecurity index, we estimate that policing alone or municipal services alone reduced insecurity by 0.05 and 0.07 standard deviations (not significant), but that insecurity fell 0.31 standard deviations in the 75 streets with both interventions.

E.3 Spillovers using an exponential rate of decay

As an alternative to estimating spillovers in our catchment areas, we can estimate a continuous, monotonic spatial decay function with the following OLS regression:

$$Y_{sqp} = \breve{\beta}_1 P_{sqp} + \breve{\beta}_2 M_{sqp} + \breve{\lambda}_1 \sum_{t \in T_P} f(d_{sqp,t}) + \breve{\lambda}_2 \sum_{t \in T_M} f(d_{sqp,t}) + \breve{\gamma}_p + \breve{\Theta} X_{sqp} + \epsilon_{sqp} \tag{3}$$

where $f(d_{sqp,t})$ is a spatial decay function with a standardized distribution. It is a weighted sum of distances to all treated hot spots, where t enumerates treated hot spots and T is the set of all treated hot spots. Treated segments receive no spillover from themselves but can receive spillovers from other treated segments. Applied to the non-experimental sample, the regression omits direct treatment effects. Our default functional form is exponential, $f(d_{sqp,t}) = 1/(e^{d_{sqp,t}})$, but we examine alternatives. We can no longer employ IPWs to weight street segments because the exposure measures are continuous variables and do not have a finite number of outcomes. Instead, we include in the control vector the expected spillover intensities (averaged across 1,000 simulations) and the probabilities of being treated by each intervention. We calculate standard errors using randomization inference.

Table E.5 reports spillovers into non-hotspots using equation (3).[8] The coefficients estimate the increase in crimes as a segment moves a standard deviation closer to a treated hot spot. The signs on the policing coefficients are all positive but not statistically significant, and consistent with our main analysis. One difference is that the evidence of displacement is no longer confined to property crimes. Here the majority of displacement seems to be associated with violent crimes. The signs on municipal services, meanwhile, are negative, implying a diffusion of benefits to nearby streets. The decrease is roughly significant at the 10% level for all crimes, and roughly significant at the 5% level for violent crimes alone. These signs are consistent across most functional forms although the statistical significance is not.

E.4 Spillovers into non-experimental non-hot spots (survey data)

Table E.6 uses our survey data estimates non-experimental spillovers within 250m using equation (2). This sample of 399 streets is too small to estimate non-experimental spillovers precisely, but the patterns are generally consistent with what we see in the large-sample dataset on reported crimes, in Table 6. The coefficients on intensive policing are positive. The coefficients on municipal services vary, but the sign on the index of overall insecurity is negative (and extremely close to zero). Unlike the effects on reported crime in the large sample, the coefficients on the interaction terms are generally negative.

F Additional robustness analysis

F.1 Aggregate effects for crime subgroups

In Tables F.2, F.1 and F.3, we display the aggregate effects on crime subgroups with confidence intervals.

[8] This functional form that places some of the heaviest weight on immediately proximate streets. Linear, logarithmic, and inverse square decay functions produce qualitatively similar conclusions, even though they give more weight to more distant segments. We ignore interactions between treatments for simplicity, as they yield similar results.

Table E.4: Impacts on insecurity, ignoring spillovers

(a) No interaction between treatments

Dependent variable	Control mean	ITT of assignment to: Intensive policing	Municipal services
	(1)	(2)	(3)
Insecurity index, z-score (+ more insecure)	0.078	-0.123	-0.160
		[.060]**	[.067]**
Perceived risk, z-score (+ riskier)	0.033	-0.116	-0.119
		[.059]**	[.065]*
Crime incidence, z-score (+ more crime)	0.096	-0.089	-0.147
		[.059]	[.068]**
Perceived & actual incidence of crime, z-score	0.039	-0.034	-0.118
		[.061]	[.072]
# crimes reported to police on street segment	1.178	-0.170	-0.164
		[.096]*	[.105]

(b) With interaction between treatments

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)
	(1)	(2)	(3)	(4)	(5)
Insecurity index, z-score (+ more insecure)	0.078	-0.049	-0.070	-0.192	-0.311
		[.055]	[.088]	[.130]	[.096]***
Perceived risk index, z-score (+ riskier)	0.033	-0.061	-0.053	-0.143	-0.257
		[.052]	[.086]	[.131]	[.096]***
Crime index, z-score (+ more crime)	0.096	-0.020	-0.065	-0.176	-0.261
		[.053]	[.089]	[.130]	[.099]***
Perceived & actual incidence of crime, z-score	0.039	-0.047	-0.134	0.033	-0.148
		[.053]	[.089]	[.139]	[.110]
# crimes reported to police on street segment	1.178	0.036	0.083	-0.530	-0.412
		[.091]	[.141]	[.202]***	[.147]***

Notes: This table reports intent to treat (ITT) estimates of the effects of the two interventions, via a WLS regression of each outcome on treatment indicators, police station (block) fixed effects, and baseline covariates (see equation 1). Panel (a) constrains the coefficient on the interaction term to be zero, and panel (b) does not. The treatment effects in panel (b) report the marginal effect of receiving any treatment or of both, and Column 5 reports the sum of the three treatment coefficients. Standard errors are clustered using the following rules: (i) for all treated segments except with cluster size 2, each segment is a cluster; (ii) for all other untreated segments, each segment gets its own cluster identifier; (iii) for entirely untreated quadrants, they form a cluster; and (iv) for quadrants with exactly 2 units assigned to treatment, those units form a cluster. The measures of perceived risk, perceived incidence of crime, and proportion reporting crime come from our citizen survey, and the # of crimes reported to the police come from police administrative data. * significant at the 10 percent, ** significant at the 5 percent, *** significant at the 1 percent.

Table E.5: Spillovers to non-hot spots with exponential rate of decay, with RI p-values

		Impact of a one standard deviation change in the average exponential distance to a hot spot treated with:	
	Control mean	Intensive policing	Municipal services
	(1)	(2)	(3)
# crimes reported to police on street segment	0.274	0.049	-0.050
		0.309	*0.102*
# property crimes only	0.100	0.004	0.001
		0.788	*0.957*
# violent crimes only	0.174	0.045	-0.051
		0.303	***0.051***

Notes: Randomization inference p-values are in italics. This table estimates the coefficients on spillovers, $\check{\lambda}$, using equation 3 above. We estimate the regression on the nonexperimental sample of non-hot spots alone. The weighted distance measures have been standardized to have zero mean and unit standard deviation.

Table E.6: Security impacts on non-hot spots <250m from treatment (N=399)

		Impact of spillovers <250m:			
Dependent variable	Control mean	Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)
	(1)	(2)	(3)	(4)	(5)
Insecurity index, z-score (+ more insecure)	-0.290	0.112	-0.002	-0.255	-0.145
		0.415	*0.966*	*0.269*	*0.435*
Perceived risk, z-score (+ riskier)	-0.099	0.018	-0.131	-0.136	-0.249
		0.925	*0.470*	*0.616*	*0.156*
Crime incidence, z-score (+ more crime)	-0.383	0.169	0.129	-0.289	0.009
		0.134	*0.372*	*0.154*	*0.822*
Perceived incidence of crime, z-score	-0.152	0.185	0.140	-0.270	0.055
		0.219	*0.478*	*0.304*	*0.741*
# crimes reported to police on street segment	0.271	0.096	0.076	-0.253	-0.081
		0.336	*0.407*	*0.167*	*0.826*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports spillover effects in the non-experimental sample from equation (2), a WLS regression of each outcome on spillover indicators, police station (block) fixed effects, and baseline covariates (1). In panel (b), Column 5 reports the sum of the three spillover coefficients.

Table F.1: Estimated aggregate property crime impacts of the interventions, accounting for spillovers <250m in the experimental and non-experimental samples

	Dependent variable: # of property crimes reported to police on segment (administrative data)							
	No interaction between treatments				Interaction between treatments			
Impacts of treatment	Coeff. (1)	RI p-value (2)	# segments (3)	Estimated total impact = (1) × (3) (4)	Coeff. (5)	RI p-value (6)	# segments (7)	Estimated total impact = (5) × (7) (8)
A. Direct treatment effect								
Intensive policing	0.006	*0.846*	756	4.2	0.087	*0.405*	756	65.7
Municipal services	-0.035	*0.997*	201	-7.1	0.094	*0.288*	201	18.9
Both					-0.343	***0.075***	75	-25.7
Subtotal				-2.8				58.8
B. Spillover, experimental sample								
Intensive policing	0.117	*0.281*	705	82.4	0.173	*0.186*	705	121.8
Municipal services	0.166	***0.043***	546	90.8	0.218	***0.030***	546	119.1
Both					-0.185	*0.356*	281	-51.8
Subtotal				173.2				189.1
C. Spillover, non-experimental sample								
Intensive policing	0.014	***0.098***	51390	708.0	0.017	***0.085***	51390	859.6
Municipal services	0.006	*0.823*	20740	126.8	0.010	*0.822*	20740	211.6
Both					-0.007	*0.442*	15491	-115.5
Subtotal				834.8				955.7
Net increase (decrease) in crime				1,005.2				1,203.6
			95% CI	(-206, 2029)			*95% CI*	(-338, 2370)
			90% CI	(-41, 1873)			*90% CI*	(27, 2203)

Notes: This table presents the aggregate effect calculation for both interventions assuming spillovers <250m. Columns 1–4 refer to the non-interacted results (equation 1 under the constraint that $\beta_3 = 0$ and $\lambda_3 = 0$) while columns 5–8 refer to the interacted results (equation 1 with no constraints). Columns 1 and 5 display the bias-adjusted treatment effect while columns 2 and 6 display RI p-values. Columns 3 and 7 display the number of units in each group. Columns 4 and 8 display the product of the bias-adjusted treatment effect and the number of units in each group. The confidence interval on the bottom of the table is constructed using randomization inference. First we create a fake schedule of potential outcomes for each observation by adding or subtracting RI-adjusted treatment or spillover effects. This process gives us a potential outcome for each unit depending on its treatment assignment. Second, we simulate a randomization and take the potential outcome associated with the treatment assignment of the new randomization. Third we estimate treatment and spillover effects using this new outcome and apply the RI bias adjustment from our main set of results. Fourth, we multiply these bias-adjusted treatment effects by the number of segments in each group, and sum across both the experimental and non-experimental samples to get the aggregate effect. We repeat steps two through four 1,000 times to get the distribution of the test statistic, which is roughly centered on the actual number of deterred crimes. The 2.5 and 97.5 percentiles of this distribution give us the 95% confidence interval. $p < .1$ in bold.

Table F.2: Estimated aggregate violent crime impacts of the interventions, accounting for spillovers <250m in the experimental and non-experimental samples

	Dependent variable: # of violent crimes reported to police on segment (administrative data)							
	No interaction between treatments				Interaction between treatments			
Impacts of treatment	Coeff. (1)	RI p-value (2)	# segments (3)	Estimated total impact = (1) × (3) (4)	Coeff. (5)	RI p-value (6)	# segments (7)	Estimated total impact = (5) × (7) (8)
A. Direct treatment effect								
Intensive policing	-0.099	***0.029***	756	-75.0	-0.077	***0.094***	756	-58.5
Municipal services	-0.041	*0.442*	201	-8.2	-0.005	*0.947*	201	-1.0
Both					-0.095	*0.214*	75	-7.1
Subtotal				-83.2				-66.6
B. Spillover, experimental sample								
Intensive policing	-0.056	*0.208*	705	-39.7	-0.030	*0.560*	705	-21.1
Municipal services	0.010	*0.700*	546	5.4	0.037	*0.358*	546	20.0
Both					-0.088	*0.211*	281	-21.1
Subtotal				-39.7				-21.1
C. Spillover, non-experimental sample								
Intensive policing	0.003	*0.449*	51190	134.1	-0.004	*0.802*	51190	-182.9
Municipal services	-0.009	***0.053***	20740	-191.5	-0.016	***0.042***	20740	-336.3
Both					0.013	*0.223*	15491	200.8
Subtotal				-57.4				-318.4
Net increase (decrease) in crime				-135.2				-365.0
			95% CI	(-858, 378)			*95% CI*	(-1112, 204)
			90% CI	(-744, 296)			*90% CI*	(-997, 82)

Notes: This table presents the aggregate effect calculation for both interventions assuming spillovers <250m. Columns 1–4 refer to the non-interacted results (equation 1 under the constraint that $\beta_3 = 0$ and $\lambda_3 = 0$) while columns 5–8 refer to the interacted results (equation 1 with no constraints). Columns 1 and 5 display the bias-adjusted treatment effect while columns 2 and 6 display RI p-values. Columns 3 and 7 display the number of units in each group. Columns 4 and 8 display the product of the bias-adjusted treatment effect and the number of units in each group. The confidence interval on the bottom of the table is constructed using randomization inference. First we create a fake schedule of potential outcomes for each observation by adding or subtracting RI-adjusted treatment or spillover effects. This process gives us a potential outcome for each unit depending on its treatment assignment. Second, we simulate a randomization and take the potential outcome associated with the treatment assignment of the new randomization. Third we estimate treatment and spillover effects using this new outcome and apply the RI bias adjustment from our main set of results. Fourth, we multiply these bias-adjusted treatment effects by the number of segments in each group, and sum across both the experimental and non-experimental samples to get the aggregate effect. We repeat steps two through four 1,000 times to get the distribution of the test statistic, which is roughly centered on the actual number of deterred crimes. The 2.5 and 97.5 percentiles of this distribution give us the 95% confidence interval. $p < .1$ in bold.

Table F.3: Estimated aggregate homicide and rape impacts of the interventions, accounting for spillovers <250m in the experimental and non-experimental samples

	Dependent variable: # of homicides and rapes reported to police on segment (administrative data)							
	No interaction between treatments				Interaction between treatments			
Impacts of treatment	Coeff. (1)	RI p-value (2)	# segments (3)	Estimated total impact = $(1) \times (3)$ (4)	Coeff. (5)	RI p-value (6)	# segments (7)	Estimated total impact = $(5) \times (7)$ (8)
A. Direct treatment effect								
Intensive policing	-0.011	*0.232*	756	-8.5	-0.008	*0.466*	756	-5.8
Municipal services	-0.006	*0.640*	201	-1.2	0.000	*0.938*	201	0.0
Both					-0.017	*0.375*	75	-1.2
Subtotal				-9.8				-7.0
B. Spillover, experimental sample								
Intensive policing	0.006	*0.643*	705	4.1	0.013	*0.360*	705	8.9
Municipal services	-0.009	*0.353*	546	-5.0	-0.002	*0.875*	546	-1.2
Both					-0.022	*0.233*	281	8.9
Subtotal				4.1				8.9
C. Spillover, non-experimental sample								
Intensive policing	-0.001	*0.419*	51190	-56.1	-0.002	*0.144*	51190	-104.5
Municipal services	0.000	*0.840*	20740	5.4	-0.001	*0.453*	20740	-16.2
Both					0.002	*0.318*	15491	32.2
Subtotal				-50.7				-88.4
Net increase (decrease) in crime				-65.4				-96.6
			95% CI	(-179, 55)			*95% CI*	(-179, 55)
			90% CI	(-165, 40)			*90% CI*	(-165, 40)

Notes: This table presents the aggregate effect calculation for both interventions assuming spillovers <250m. Columns 1–4 refer to the non-interacted results (equation 1 under the constraint that $\beta_3 = 0$ and $\lambda_3 = 0$) while columns 5–8 refer to the interacted results (equation (1) with no constraints). Columns 1 and 5 display the bias-adjusted treatment effect while columns 2 and 6 display RI p-values. Columns 3 and 7 display the number of units in each group. Columns 4 and 8 display the product of the bias-adjusted treatment effect and the number of units in each group. The confidence interval on the bottom of the table is constructed using randomization inference. First we create a fake schedule of potential outcomes for each observation by adding or subtracting RI-adjusted treatment or spillover effects. This process gives us a potential outcome for each unit depending on its treatment assignment. Second, we simulate a randomization and take the potential outcome associated with the treatment assignment of the new randomization. Third we estimate treatment and spillover effects using this new outcome and apply the RI bias adjustment from our main set of results. Fourth, we multiply these bias-adjusted treatment effects by the number of segments in each group, and sum across both the experimental and non-experimental samples to get the aggregate effect. We repeat steps two through four 1,000 times to get the distribution of the test statistic, which is roughly centered on the actual number of deterred crimes. The 2.5 and 97.5 percentiles of this distribution give us the 95% confidence interval. $p < .1$ in bold.

Table F.4: Aggregate impacts on crimes by type, assuming two levels of spillovers (mean and confidence intervals)

	without interaction			*with interaction*		
	Effect	95% CI	90% CI	Effect	95% CI	90% CI
	(1)	(2)	(3)	(4)	(5)	(6)
All crime	796.0	(-1548, 3565)	(-1157, 3149)	1000.1	(-1710, 3590)	(-1224, 3198)
Property crime	374.9	(-1484, 2563)	(-1211, 2315)	620.8	(-1542, 2673)	(-1144, 2413)
Violent crime	502.5	(-750, 1608)	(-510, 1431)	450.7	(-854, 1503)	(-670, 1375)
Homicides and sexual assaults only	70.9	(-188, 251)	(-150, 215)	63.3	(-210, 241)	(-162, 211)

Notes: This table presents the aggregate effect calculation for various crime subgroups assuming spillovers at <250m and 250-500m simultaneously. Calculations are based on the aggregate effect and confidence interval described in Table 6.

F.2 Aggregate effects assuming two levels of spillovers

In Table F.4 we summarize aggregate impacts on crimes by type, assuming two levels of spillovers simultaneously: <250m and 250-500m. This is an alternative version of equation (1) with separate dummies for streets located within 250m and streets located 250-500m from treatment hot spots.

F.3 Impacts without re-weighting and randomization inference

Table F.5 reproduces Table 8 from the paper, but without IPWs and randomization inference. The direct treatment effects are generally smaller but the patterns are still similar. However, the spillover effects in these results are huge (.18 standard deviations for hot spots policing, 0.31 standard deviations for municipal services). This shows that IPW's are crucial for getting the spillover effects right– the point estimates on the direct effects do not change as much because most segments have similar probabilities of being treated.

Thus estimating unbiased treatment and spillover effects in the presence of the geographic clustering of high crime areas requires the use of inverse probability weights and randomization inference.

F.4 Impacts over time

With daily administrative crime reports, we can calculate cumulative treatment effects with each day or week of the interventions. Figure F.1 reports the results of estimating equation (1) on the cumulative level of reported crime starting 5 weeks after intensive policing began, and continuing to the end of the intervention period. The cumulative effect of both interventions appears within 8 to 12 weeks of the intervention, and grows over time. In particular, we see that the marginal effect of receiving both interventions grows larger as time passes.

F.5 Disentangling municipal services

First, we see no evidence that municipal services treatment effects were concentrated in the hot spots diagnosed as needing improved lights. Table F.6 reports ITT effects of both interventions (without an interaction) for the full sample (column 1) and also for the subsample of 414 hot spots where the city conducted a lights needs assessment (columns 2–4).[9] The coefficient on the municipal services intervention is closer to zero in the lights eligible case (column 3) and is not statistically significant for the clean-up only (lights ineligible) hot spots.

Second, we don't see larger treatment effects at nighttime. We use the recorded time of a crime in police

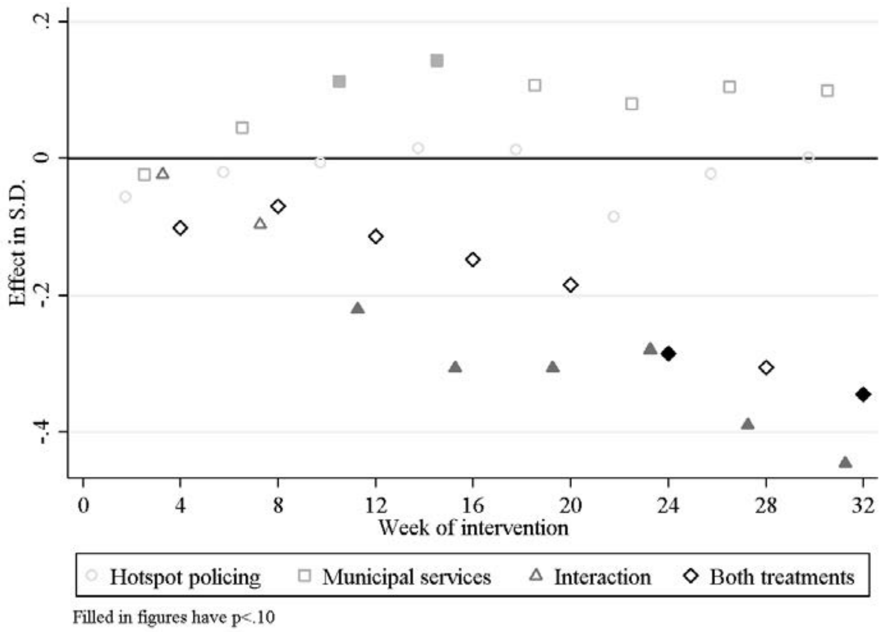
[9] Recall that 201 streets were assigned to be eligible for the municipal services treatment. At the same time we selected an additional 213 for assessment, in order to be able to have baseline data on this lights needs assessment for this analysis.

Table F.5: Naïve treatment effects

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interven- tions	Sum of (2), (3), and (4)	Impact of spillovers <250: Any intensive policing	Any municipal services	Both interven- tions	Sum of (6), (6), and (8)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Insecurity index, z-score (+ more insecure)	0.066	0.006	-0.069	-0.090	-0.153	0.180	0.308	-0.225	0.262
		[.057]	[.091]	[.130]	[.102]	[.064]***	[.066]***	[.096]**	[.078]***
Perceived risk index, z-score (+ riskier)	-0.018	-0.022	-0.080	-0.096	-0.198	0.107	0.215	-0.193	0.129
		[.058]	[.088]	[.131]	[.104]*	[.063]*	[.064]***	[.088]**	[.076]*
Crime index, z-score (+ more crime)	0.128	0.031	-0.035	-0.053	-0.058	0.192	0.297	-0.182	0.308
		[.055]	[.095]	[.134]	[.103]	[.067]***	[.067]***	[.105]*	[.084]***
Perceived & actual incidence of crime, z-score	0.002	-0.002	-0.120	0.101	-0.021	0.143	0.240	-0.211	0.172
		[.062]	[.090]	[.143]	[.120]	[.067]**	[.070]***	[.097]**	[.082]**
# crimes reported to police on street segment	1.334	0.086	0.135	-0.333	-0.113	0.238	0.334	-0.082	0.491
		[.091]	[.160]	[.217]	[.164]	[.124]*	[.115]***	[.197]	[.164]***

Notes: This table reports intent to treat (ITT) estimates of the effects of the two interventions, via a OLS regression of each outcome on treatment indicators, police station (block) fixed effects, and baseline covariates (see equation (1). Column 5 reports the sum of the three treatment coefficients. Standard errors are clustered using the following rules: For all treated segments except with cluster size 2, each segment is a cluster. For all other untreated segments, each segment gets its own cluster ID. For entirely untreated quadrants, they form a cluster. For quadrants with exactly 2 units assigned to treatment, those units form a cluster.

Figure F.1: Impacts on reported crime over time, in weeks since treatment began (administrative crime data only)



Notes: The figure reports the ITT effect of the two interventions and the interaction term, plus the sum of these three coefficients. The sample accumulates the number of weeks of administrative data on crime reports included, starting five weeks after the intensive policing treatment began.

Table F.6: Municipal services impacts by subgroup

Independent variable	Dependent variable: Index of insecurity (z-score)			
		Block 1 versus Block 2		
	Full sample	All	Lights eligible only	Lights ineligible only
	(1)	(2)	(3)	(4)
Assigned to intensive policing	-0.095	-0.132	-0.430	-0.133
	[0.075]	[0.124]	[0.322]	[0.131]
Assigned to municipal services	-0.096	-0.010	0.200	-0.043
	[0.074]	[0.105]	[0.317]	[0.131]
<250m from any unit assigned to intensive policing	0.050	0.195	-0.043	0.144
	[0.076]	[0.140]	[0.333]	[0.139]
<250m from any unit assigned to municipal services	0.164	0.258	0.689	0.221
	[0.061]***	[0.165]	[0.275]**	[0.196]
Number of observations	1,916	414	120	294

Notes: This table reports the same intent to treat (ITT) estimates on the insecurity index as in Table E.4a (Column 1) and the same analysis in three subsamples: all 414 segments assigned to Block 1 or 2 of municipal services treatment that received a city assessment (Column 2); the 120 segments in Blocks 1 and 2 that were deemed eligible for lighting improvement (Column 3); and the 294 segments that were not (Column 4). * significant at the 10 percent, ** significant at the 5 percent, *** significant at the 1 percent.

Table F.7: Impacts on insecurity by time of day

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)	Impact of spillover <250m: Any intensive policing	Any municipal services	Both interventions	Sum of (1), (2), and (3)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Perceived risk index, z-score	0.049	-0.122	-0.086	-0.084	-0.292	0.002	0.083	-0.160	-0.075
		0.259	*0.494*	*0.644*	*0.094*	*0.511*	*0.129*	***0.085***	*0.808*
Components related to daytime risk	0.019	-0.111	0.005	-0.164	-0.270	-0.022	0.091	-0.122	-0.054
		0.321	*0.736*	*0.277*	*0.139*	*0.642*	*0.103*	*0.185*	*0.723*
Components related to risk after dark	0.069	-0.122	-0.132	-0.043	-0.297	0.016	0.066	-0.179	-0.097
		0.265	*0.245*	*0.887*	***0.098***	*0.451*	*0.191*	***0.070***	*0.896*
# crimes reported to police	0.743	0.009	0.089	-0.437	-0.339	0.143	0.255	-0.272	0.125
		0.817	*0.367*	***0.043***	*0.109*	*0.315*	***0.025***	*0.196*	*0.289*
Daytime (6 a.m. – 6 p.m.)	0.472	-0.054	0.026	-0.148	-0.176	0.071	0.232	-0.246	0.058
		0.568	*0.666*	*0.337*	*0.172*	*0.578*	***0.004***	***0.084***	*0.570*
Nighttime (6 p.m. – 6 a.m.)	0.271	0.064	0.063	-0.289	-0.163	0.071	0.022	-0.026	0.067
		0.255	*0.290*	***0.005***	*0.260*	*0.273*	*0.534*	*0.895*	*0.262*

Notes: p-values generated via randomization inference are in italics, with $p < .1$ in bold. This table reports intent to treat (ITT) estimates of the effects of the two interventions, via a WLS regression of each outcome on treatment indicators, police station (block) fixed effects, and baseline covariates (see equation 1). Column 5 reports the sum of the three treatment coefficients.

data, and divide perceived risk questions into those that relate to nighttime and daytime risk.[10] Table F.7 reports results with the interaction. In general, the coefficients on nighttime and daytime risk have the same sign and magnitude, especially in the 75 streets where both treatments (and treatment effects) were concentrated. (Note that control risk at daytime is smaller than during the night, so a coefficient of similar size turns out to be more important at daytime that nighttime.) Thus there is no evidence treatment effects are greater in nighttime.

References

Braga, A. and B. J. Bond (2008). Policing crime and disorder hot spots: A randomized controlled trial. *Criminology 46*, 577–608.

Braga, A., A. V. Papachristos, and D. M. Hurreau (2012). An ex post factor evaluation framework for place-based police interventions. *Campbell Systematic Reviews 8*, 1–31.

Braga, A., D. Weisburd, E. Waring, L. Green Mazerolle, W. Spelman, and F. Gajewski (1999). Problem-oriented policing in violent crime places: A randomized controlled experiment. *Criminology 37*, 541–580.

Chalfin, A. and J. McCrary (forthcoming 2017b). Are US Cities Underpoliced?: Theory and Evidence. *Review of Economics and Statistics*.

Collazos, D., E. Garcia, D. Mejia, D. Ortega, and S. Tobon (2017). Hotspots policing in a high crime environment: An experimental evaluation in Medellin. *In progress*.

Di Tella, R. and E. Schargrodsky (2004, March). Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack. *American Economic Review 94*(1), 115–133.

Gerber, A. S. and D. P. Green (2012). *Field experiments: Design, analysis, and interpretation*. New York: WW Norton.

Mazerolle, L. G., J. F. Prince, and J. Roehl (2000). Civil remedies and drug control: a randomized field trial in Oakland, CA. *Evaluation Review 24*, 212–241.

Ratcliffe, J. H., T. Tangiguchi, E. R. Groff, and J. D. Wood (2011). The Philadelphia Foot Patrol Experiment: A Randomized Controlled Trial of Police Patrol Effectiveness in Violent Crime Hotspots. *Criminology 49*, 795–831.

Sherman, L., M. Buerger, and P. Gartin (1989). Beyond dial-a-cop: A randomized test of repeat call policing (recap). Washington, D.C.: Crime Control Institute.

Sherman, L. and D. P. Rogan (1995). Deterrent effects of police raids on crack houses: A randomized, controlled experiment. *Justice Quarterly 12*(4), 755–781.

Sherman, L. and D. Weisburd (1995). Does Patrol Prevent Crime? The Minneapolish Hot Spots Experiment. In *Crime Prevention in the Urban Community*. Boston: Kluwer Law and Taxation Publishers.

Taylor, B., C. Koper, and D. Woods (2011). A randomized controlled trial of different policing strategies at hot spots of violent crime. *Journal of Experimental Criminology 7*, 149–181.

Telep, C., R. Mitchell, and D. Weisburd (2014). How Much Time Should Police Spend at Crime Hot Spots? Answers from a Police Agency Directed Randomized Field Trial in Sacramento, California. *Justice Quarterly 31*(5), 905–933.

Weisburd, D. and C. Gill (2014). Block Randomized Trials at Places: Rethinking the Limitations of Small N Experiments. *Journal of Quantitative Criminology 30*(1), 97–112.

Weisburd, D. and L. Green (1995). Measuring Immediate Spatial Displacement: Methodological Issues and Problems. In *Crime and Place: Crime Prevention Studies*, pp. 349–359. Monsey, NY: Willow Tree Press.

[10] The city is near the equator and so 6 p.m. to 6 a.m. roughly corresponds to dusk, dark and dawn year round. Nighttime risk questions include general risk at dusk, for a young woman to walk alone after dark, and for a young man to walk alone after dark. Daytime questions include general daytime risk and risk of talking on one's smartphone.